



# Etude à grande échelle du rôle de TFIIS et de ses partenaires dans la transcription chez les eucaryotes

Yad Ghavi-Helm

## ► To cite this version:

Yad Ghavi-Helm. Etude à grande échelle du rôle de TFIIS et de ses partenaires dans la transcription chez les eucaryotes. Biochimie [q-bio.BM]. Université Paris Sud - Paris XI, 2009. Français. NNT : . tel-00446842

**HAL Id: tel-00446842**

**<https://theses.hal.science/tel-00446842>**

Submitted on 13 Jan 2010

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITE DE PARIS-SUD U.F.R. SCIENTIFIQUE D'ORSAY

THESE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE PARIS XI ORSAY

École doctorale "Gènes, Génomes, Cellules"

Présentée et soutenue publiquement le 25 septembre 2009

par

Yad Ghavi-Helm

Etude à grande échelle du rôle de TFIIS et de ses  
partenaires dans la transcription chez les eucaryotes.

Jury

Pierre Capy, Président

Nouria Hernandez, Rapporteur

Bertrand Séraphin, Rapporteur

Alain Jacquier, Examineur

Pierre Thuriaux, Examineur

Julie Soutourina, Directrice de thèse



*Ce n'est pas dans la science qu'est le bonheur,  
mais dans l'acquisition de la science*

*Edgar Allan Poe*





# Remerciements

*Eh voilà! Trois années se sont (trop) vite écoulées, et arrive déjà le moment de faire le bilan de cette expérience unique qu'est une thèse et de remercier les nombreuses personnes sans qui ces années n'auraient pas été ce qu'elles ont été.*

*Je tiens tout d'abord à remercier Monsieur Pierre Capy, Madame Nouria Hernandez, Monsieur Bertrand Séraphin, Monsieur Alain Jacquier et Monsieur Pierre Thuriaux d'avoir accepté d'être membres de mon jury de thèse.*

*Je remercie tout particulièrement Pierre et Michel pour m'avoir accueillie au sein du laboratoire. Merci Pierre pour ta disponibilité et ta passion communicante pour la recherche. Merci Michel de t'être toujours intéressé à mon travail et de m'avoir toujours permis de disposer des moyens nécessaires à mon travail de recherche. Merci à tous les deux d'avoir animé nos "discussions à 4", parfois mouvementées, mais toujours enrichissantes.*

*Un grand merci à Julie, qui a su me guider de façon subtile tout au long de ma thèse, tout en me laissant toute l'indépendance dont j'avais besoin pour m'épanouir. Peu de gens peuvent se vanter de pouvoir entrer à tout moment dans le bureau de leur directeur de thèse pour discuter de science (ou pas!), ça a été mon cas, et je t'en remercie. Tu resteras toujours pour moi un modèle de rigueur, d'énergie et d'efficacité.*

*Je remercie également tous les membres de l'équipe Werner/Thuriaux que j'ai pu cotoyer au cours de ces années. Cyril, Fred, Sylvain, l'inoubliable chorale du labo (Oh Marie...si tu savais). Merci Cyril pour ta bonne humeur communicative qui m'a bien aidée à ne pas m'arracher trop de cheveux devant mes tableaux excel! Merci aux "Thuriaux's girls", Marta, Léna, Daphné pour vos conseils avisés. Sylvie, je n'oublierais pas ton énergie, ton amour de la belle génétique, et surtout tes délicieux macarons! Merci Claire pour ta présence discrète mais indispensable à la vie du labo, surtout lorsque les pièces 52/54 se sont retrouvées bien vides...*

*Au cours de la deuxième moitié de ma thèse, de nouvelles têtes sont arrivées, prêtes à prendre la relève. Sébastien, merci de m'avoir initiée à la technique du Recombineering. Lucie, merci de m'avoir transmis tes connaissances en matière de culture des cellules ES, de tes conseils et de ta grande gentillesse. Je te souhaite le meilleur pour ta fin de thèse et pour la*

suite! Sandra et Marina, votre arrivée est venue agayer ma dernière année de thèse. Merci pour votre sympathie et d'avoir supporté ma mauvaise humeur pendant l'écriture de cette thèse.

Je tiens également à remercier l'ensemble des membres de l'équipe de Matthieu Gérard, sans qui je n'aurais pas pu découvrir les joies de la culture cellulaire! Merci Matthieu, Maud, Isabelle, Arnaud, Faycal de m'avoir fait une petite place sous les hottes. Merci Hélène et Sylvie pour vos conseils et votre travail indispensable au fonctionnement du labo. Merci également à tous les membres de l'animalerie pour leur gentillesse et leur accueil.

Une grande partie de ce travail n'aurait pas pu voir le jour sans l'aide des "biochimistes du 1<sup>er</sup> étage". Merci Joël de m'avoir initiée à la transcription *in vitro* et à la purification de protéines, mais également de m'avoir laissé utiliser tes précieux aliquots de protéines! Merci à Audrey et Christine de m'avoir révélé les subtilités des puces Pol III. Nayla, merci pour tes conseils pour la purification des complexes ternaires et de ton soutien au cours des longues heures passées dans la chambre froide. Merci Olivier, Joël et Christine vos conseils et les discussions que nous avons pu avoir autour de la Pol III.

Merci à Magali et Jean-Christophe de m'avoir initiée au monde merveilleux de la bioinformatique. Merci Magali pour ta patience et tes précieux conseils.

Merci à tous les participants des fameuses "bier session" du vendredi soir. Je n'ai pas toujours eu le temps de venir, mais j'ai toujours passé de bons moments en votre compagnie (et appris tous les derniers potins du SBTGeM!).

Merci également aux personnes incontournables du bâtiment 144, Arlette "ultra efficace", Manu "faut pas commander la veille des vacances", Catherine "je vais essayer de faire passer ta demande de HNO pour dimanche", Chantal "toujours souriante", et Linda "la perle du Portugal".

Enfin, je tiens à remercier mes amis et ma famille, qui ont eu le courage de me supporter pendant ces trois ans, malgré mes sautes d'humeur et mes "j'ai pas le temps" et "je dois aller au labo ce week-end". Je remercie tout particulièrement mes parents d'avoir toujours su me prodiguer de bons (même si je ne l'ai pas toujours admis) conseils dans les moments difficiles.

# Table des matières

<b>AVANT PROPOS</b> .....	7
<b>INTRODUCTION</b> .....	9
<b>CHAPITRE I</b> .....	11
<b>La transcription chez les eucaryotes</b> .....	11
a) Transcription par l'ARN polymérase II .....	12
<u>L'ARN polymérase II: structure et fonction</u> .....	12
<u>Le cycle de transcription par l'ARN polymérase II (Figure 2)</u> .....	15
<u>Le facteur général de transcription TFIID</u> .....	22
<u>La régulation de la transcription par l'ARN polymérase II</u> .....	24
b) Transcription par l'ARN polymérase III.....	30
<b>CHAPITRE II</b> .....	35
<b>Le facteur de transcription TFIIS</b> .....	35
a) Identification de TFIIS et de ses activités <i>in vitro</i> .....	35
b) Structure et mode d'action de TFIIS <i>in vitro</i> .....	38
c) Fonction de TFIIS .....	41
d) Le clivage du transcrit : une fonction conservée .....	42
e) Rôle de TFIIS dans l'initiation de la transcription .....	46
<b>CHAPITRE III</b> .....	48
<b>Exploration fonctionnelle des génomes: de la levure à l'homme</b> .....	48
a) Le séquençage des génomes eucaryotes: utilité de <i>S. cerevisiae</i> comme modèle.....	48
b) Les puces à ADN: présentation générale de la technique.....	50
c) Les études pionnières du transcriptome chez <i>S. cerevisiae</i> .....	55
b) Distribution des protéines sur le génome: immunoprécipitation de la chromatine.....	59
<u>Cibles des régulateurs transcriptionnels</u> .....	61
<u>Distribution des histones et de leurs modifications</u> .....	61
<u>Distribution des ARN polymérases et de leurs facteurs généraux</u> .....	63
c) Apports récents du séquençage à haut-débit.....	67
<b>RESULTATS ET DISCUSSION</b> .....	75
<b>I/ Etude de la localisation globale du facteur TFIIS</b> .....	77
a) Rôle de TFIIS dans la transcription par l'ARN polymérase III chez <i>S. cerevisiae</i> .....	77
<u>Contexte du travail</u> .....	77
<u>Article</u> .....	81
b) Rôle des homologues murins de TFIIS dans la régulation de la transcription chez <i>Mus musculus</i> .....	121
<u>Contexte du travail</u> .....	121
<u>Résultats</u> .....	122
<b>II/ Rôle du Médiateur dans le recrutement du complexe TFIID lors de la formation du PIC in vivo</b> .....	131
<u>Contexte du travail</u> .....	131
<u>Article</u> .....	134
<b>III/ L'ARN polymérase II: senseur de la disponibilité en nucléosides triphosphates</b> .....	169
<u>Contexte du travail</u> .....	169
<u>Article</u> .....	171
<b>CONCLUSIONS ET PERSPECTIVES</b> .....	193
<u>Rôle de TFIIS dans la transcription par l'ARN polymérase II</u> .....	195
<u>TFIIS et Bye1: un rôle antagoniste dans la transcription par l'ARN polymérase II?</u> .....	196
<u>Rôle de TFIIS dans la transcription par l'ARN polymérase III</u> .....	198

<u>TFIIS : un facteur commun aux trois machineries de transcription eucaryotiques?</u> .....	201
<u>Rôle des différentes isoformes de TFIIS chez la souris</u> .....	203
<u>Rôle du Médiateur dans la mise en place du PIC</u> .....	204
<u>Etude de mutants de l'ARN polymérase II: un outil pour la compréhension des mécanismes de la transcription par l'ARN polymérase II</u> .....	205
<b><i>BIBLIOGRAPHIE</i></b> .....	<b>207</b>
<b><i>ANNEXES</i></b> .....	<b>233</b>

# Table des illustrations

Figure 1. Structure de l'ARN polymérase II obtenue par diffraction des rayons X (d'après Armache et al., 2003).	13
Figure 2. Les étapes de la transcription (d'après Margaritis & Holstege, 2008).	17
Figure 3. ARN polymérase II en situation de pause au promoteur et transition vers l'étape d'élongation (d'après Wu & Snyder, 2008).	19
Figure 4. L'initiation et la réinitiation de la transcription par l'ARN polymérase II (d'après Hahn, 2004).	21
Figure 5. Modèle de microscopie électronique des complexes TFIID humain et de levure (d'après Chang & Kornberg, 2000; Schultz et al., 2000).	23
Figure 6. Organisation du Médiateur (d'après Guglielmi et al., 2004).	26
Figure 7. Interaction du Médiateur avec l'ARN polymérase II (d'après Chadick & Asturias, 2005).	27
Figure 8. Carte d'interactions génétiques du Médiateur chez <i>S. cerevisiae</i> (d'après Collins et al., 2007).	27
Figure 9. Modèle d'organisation spatiale de l'ARN polymérase III de <i>S. cerevisiae</i> (d'après Flores et al., 1999; Lorenzen et al., 2007).	30
Figure 10. Les différents types de promoteurs de l'ARN polymérase III (d'après Schramm & Hernandez, 2002).	32
Figure 11. Recrutement de la machinerie de transcription par l'ARN polymérase III (d'après Schramm & Hernandez, 2002).	32
Figure 12. Modèle du passage par l'ARN polymérase II des sites d'arrêt grâce à l'activité de clivage de TFIIS (d'après Cramer, 2004).	36
Figure 13. Architecture de TFIIS (d'après Kettenberger et al., 2003).	39
Figure 14. Alignement des séquences des protéines TFIIS de différents organismes (d'après Fish & Kane, 2002).	44
Figure 15. Conservation des domaines de TFIIS chez les protéines Rpa12 (ARN polymérase I), Rpc11 (ARN polymérase III) et Rpb9 (ARN polymérase II).	46
Figure 16. Schéma de la synthèse des oligonucléotides chez Agilent (d'après <a href="http://www.chem.agilent.com/en-US/Products/Instruments/dnamicarrays/Pages/gp557.aspx">http://www.chem.agilent.com/en-US/Products/Instruments/dnamicarrays/Pages/gp557.aspx</a> ).	53
Figure 17. Les micro-miroirs utilisés par Nimblegen (d'après <a href="http://www.nimblegen.com/technology/manufacture.html">http://www.nimblegen.com/technology/manufacture.html</a> ).	53
Figure 18. Schéma de synthèse <i>in situ</i> des oligonucléotides chez Nimblegen. (d'après <a href="http://www.nimblegen.com/technology/manufacture.html">http://www.nimblegen.com/technology/manufacture.html</a> ).	54
Figure 19. Schéma de synthèse <i>in situ</i> des oligonucléotides chez Affymetrix (d'après Lipshutz et al., 1999).	55
Figure 20. Résumé de la technique de ChIP-on-chip (d'après Buck & Lieb, 2004).	60
Figure 21. Vue schématique de l'architecture de la chromatine chez <i>S. cerevisiae</i> (d'après Liu et al., 2005).	63
Figure 22. Comparaison du nombre de transcrits par heure au ratio d'enrichissement moyen de l'ARN polymérase II (d'après Steinmetz et al., 2006).	65
Figure 23. Séquençage par la méthode de Sanger.	69
Figure 24. Les technologies de séquençage à haut débit (d'après Medini et al., 2008).	72
Figure 25. Organisation chromosomique des gènes <i>Tcea1</i> , <i>Tcea2</i> et <i>Tcea3</i> .	122
Figure 26. Schéma de la technique d'étiquetage de protéines par "Recombineering".	124
Figure 27. Schéma de la construction de lignées cellulaires étiquetées pour <i>Tcea1</i> , <i>Tcea2</i> et <i>Tcea3</i> .	125
Figure 28. Génotypage des clones de <i>Tcea1</i> étiquetés par Southern blot.	127
Figure 29. Tailles des bandes attendues lors du génotypage des lignées cellulaires étiquetées pour <i>Tcea1</i> , <i>Tcea2</i> ou <i>Tcea3</i> .	127
Figure 30. Observation du caryotype d'un clone positif <i>Tcea1</i> étiqueté.	128
Figure 31. Expression des clones positifs <i>Tcea1</i> étiquetés.	128
Figure 32. Alignement de la séquence des domaines de liaison à l'ADN de TFIIS et Bye1 chez <i>S. cerevisiae</i> .	197
Figure 33. Fonctions de TFIIS dans la transcription.	203

Tableau 1: Les facteurs généraux de la transcription. ....	16
Tableau 2: Le complexe TFIIH chez l'homme et la levure <i>S. cerevisiae</i> . ....	22
Tableau 3: Sous-unités du Médiateur chez <i>S. cerevisiae</i> (d'après Bourbon, 2008). ....	29
Tableau 4: Transcrits de l'ARN polymérase III chez <i>S. cerevisiae</i> . ....	31
Tableau 5: Les facteurs généraux de transcription par l'ARN polymérase III chez <i>S. cerevisiae</i> . ....	33
Tableau 6: Oligonucléotides utilisés pour la construction des vecteurs de recombinaison. ....	126

## Abréviations

6AU: 6-azauracil

ADN: Acide DésoxyriboNucléique

ARN: Acide RiboNucléique

ARNm: ARN messenger

ARNr: ARN ribosomal

ARNt: ARN de transfert

ATP: Adénosine TriPhosphate

BRE: Élément de reconnaissance par TFIIB ("TFIIB Recognition Element")

ChIP: Immunoprécipitation de chromatine ("Chromatin ImmunoPrecipitation")

CTD: Domaine carboxy-terminal de Rpb1 ("Carboxy Terminal Domain")

CGH: Hybridation génomique comparative ("Comparative Genomic Hybridization")

CoIP: Co-ImmunoPrécipitation

CUT: Transcrit cryptique instable ("Cryptic Unstable Transcript")

dNTP: désoxyriboNucléoside TriPhosphate

ddNTP: didésoxyriboNucléoside TriPhosphate

DMD: matrice de micro-miroirs ("Digital Micromirror Device")

DPE: Élément du promoteur aval ("Downstream Promoter Element")

DSE: élément de séquence distal ("distal sequence element")

DSIF: "DRB Sensitivity-Inducing Factor"

ELL: "eleven-nineteen lysine-rich in leukaemia"

ES: cellule souche embryonnaire ("Embryonic Stem cell")

ESR: "enviromental stress response"

FACT: "Facilitates Chromatin Transcription"

GFP: "Green Fluorescent Protein"

GTF: Facteur Général de Transcription

GTP: Guanosine TriPhosphate

HA: "Human influenza hemagglutinin"

HAT: Histone Acetyl-Transférase

HDAC: histone déacétylase

ICR: région de contrôle interne ("internal control region")

IE: élément intermédiaire ("intermediate element")



INR: Séquence initiatrice ("Initiator")  
LM-PCR: "ligation-mediated-PCR"  
MAS: "Maskless Array Synthesizer"  
MEF: fibroblaste embryonnaire murin ("Mouse embryonic fibroblast")  
MPA: Mycophénolate  
NELF: "negative elongation factor"  
NER: réparation par excision de nucléotides ("nucleotide excision repair")  
NFR: région dépourvue de nucléosomes ("nucleosome-free promoter region")  
NTP: Nucléoside TriPhosphate  
ORF: Phase ouverte de lecture ("Open Reading Frame")  
PCR: "Polymerase Chain Reaction"  
PIC: Complexe de préinitiation ("PreInitiation Complex")  
PSE: élément de séquence proximal ("proximal sequence element")  
P-TEFb: "Positive Transcription-Elongation Factor-b"  
RMN: résonance magnétique nucléaire  
RT-PCR: "Reverse Transcription Polymerase Chain Reaction"  
SAGA: "Spt-Ada-Gnc5 Acetyltransferase"  
SINE: "Short Interspersed repeated DNA elements"  
SNP: "Single Nucleotide Polymorphism"  
TAF: Facteurs associés à TBP "TBP Associated Factors"  
TAP: "Tandem Affinity Purification"  
TBP: Protéine de liaison à la boîte TATA "TATA-Binding Protein"  
TRAMP: "Trf4-Air1/Air2-Mtr4 polyadenylation complex"  
TSS: "Transcription Start Site"  
UAS: "Upstream Activating Sequence"  
UTP: Uridine TriPhosphate  
UV UltraViolet  
WT: type sauvage ("Wild Type")  
XP: *xeroderma pigmentosum*

## AVANT PROPOS

A mon arrivée au laboratoire, la technologie des puces à ADN amorçait un tournant avec le développement des puces "à haute résolution". J'ai ainsi été impliquée dans la mise en place des protocoles d'hybridation sur les puces commerciales Agilent pour les expériences de localisation à grande échelle par ChIP-on-chip. Pendant ma thèse, les nouvelles plateformes de séquence à haut débit ont fait leur apparition, ouvrant ainsi de nouvelles perspectives pour l'analyse de la localisation globale des protéines sur le génome par ChIP-Seq.

Au cours de ma première année de thèse, j'ai utilisé le ChIP-on-chip pour l'analyse de la localisation globale du facteur de transcription TFIIS et de l'ARN polymérase II sur le génome de *Saccharomyces cerevisiae*. TFIIS est le premier facteur de transcription de l'ARN polymérase II à avoir été identifié et a, depuis lors, été largement étudié. Ces études ont permis de montrer le rôle de TFIIS dans la stimulation de l'activité de clivage intrinsèque de l'ARN polymérase II. En effet, lorsque l'ARN polymérase II rencontre un site de pause, celle-ci subit un mouvement de recul, entraînant un décalage entre le site actif de l'enzyme et l'extrémité 3' de l'ARN en cours de synthèse. L'ARN polymérase II est alors incapable de repartir en élongation tant que l'ARN extrudé n'est pas clivé. Peu de temps avant mon arrivée, plusieurs équipes, dont la nôtre, ont mis en évidence un second rôle de TFIIS dans l'initiation de la transcription par l'ARN polymérase II, qui serait indépendant de son activité de clivage. TFIIS agirait, en conjonction avec le Médiateur, dans la mise en place du complexe de préinitiation.

La problématique initiale visait donc à mieux comprendre le rôle de TFIIS *in vivo*, et notamment de déterminer si l'analyse de la localisation globale de l'ARN polymérase II permet de déterminer la position des sites de pauses *in vivo*, si TFIIS est localisé uniquement sur ces sites de pauses, et sur quels promoteurs et quels gènes TFIIS est localisé. A notre grande surprise, en plus de nous permettre de répondre, dans une certaine mesure, à ces interrogations, nos expériences de localisation à grande échelle ont révélé la présence de TFIIS sur la quasi-totalité des gènes transcrits par l'ARN polymérase III. Par la suite, nous avons approfondi cette observation et montré, à la fois *in vivo* et *in vitro*, que TFIIS est un facteur de transcription de l'ARN polymérase III.

Au cours de ma dernière année de thèse, je me suis attachée à élargir ces résultats en étudiant le rôle de TFIIS chez la souris. Pour ce faire, nous avons envisagé, dans un premier

temps, de déterminer la localisation globale des trois isoformes murines de TFIIS dans des cellules souches embryonnaires de souris par des expériences de ChIP-Seq. Une part importante du projet a consisté à tenter d'obtenir des lignées de cellules souches étiquetées pour chacune de ces trois isoformes. Les premières expériences d'immunoprécipitation de la chromatine sont actuellement en cours au laboratoire.

En parallèle, j'ai également été impliquée dans deux projets portant sur des partenaires de TFIIS. J'ai ainsi, en collaboration avec Cyril Esnault au laboratoire, participé à l'étude du rôle du Médiateur dans la mise en place du complexe de préinitiation *in vivo*. J'ai également étudié la localisation globale de l'ARN polymérase II, à la fois dans une souche sauvage et dans différents mutants de cette enzyme. L'analyse de la localisation globale de l'ARN polymérase II dans le mutant *rpb1-L1397S* a ainsi contribué à la mise en évidence d'un rôle de l'ARN polymérase II comme senseur de la disponibilité en nucléosides triphosphates dans la cellule.

La première partie de ce manuscrit présentera les connaissances actuelles sur la transcription par les ARN polymérases II et III et sur le facteur TFIIS. Je reviendrai ensuite sur l'évolution de l'analyse fonctionnelle des génomes en m'appuyant sur différents exemples, notamment chez *S. cerevisiae*. Je décrirai ensuite les résultats obtenus au cours de ma thèse sous forme de trois publications scientifiques et de résultats préliminaires pour l'étude de TFIIS chez la souris. Enfin, ces travaux seront mis en perspectives, de manière à mettre en avant les élargissements possibles pour l'étude du facteur TFIIS et de ses partenaires.

# **INTRODUCTION**



## **CHAPITRE I**

### ***La transcription chez les eucaryotes***

La transcription est un processus essentiel, partagé par l'ensemble des organismes vivants et la première étape pour l'expression des gènes. Sa régulation est finement contrôlée et met en œuvre de nombreux facteurs ou complexes multiprotéiques. Chez les eucaryotes, la transcription nucléaire est assurée par les ARN polymérases I, II et III. L'ARN polymérase I est responsable de la transcription du précurseur des grands ARN ribosomiaux (ARNr), alors que l'ARN polymérase II transcrit majoritairement les ARN messagers (ARNm), ainsi que certains ARN non codants. Le nombre d'ARN non codants connu s'est considérablement accru ces dernières années, notamment grâce à l'émergence des techniques de séquençage à haut débit (voir aussi Chapitre III). Enfin, l'ARN polymérase III synthétise les ARN de transfert (ARNt), l'ARN ribosomal 5S ainsi que d'autres petits ARN stables et non traduits. Deux ARN polymérases supplémentaires, appelées ARN polymérases IV et V, ont également été identifiées chez les plantes. Elles sont impliquées dans la production des ARNs du silencing, mais ne semblent pas avoir d'équivalent chez les autres eucaryotes.

Le cycle de transcription des ARN polymérases peut être divisé en trois étapes principales: l'initiation, au cours de laquelle l'ARN polymérase et ses facteurs d'initiation sont recrutés et permettent la formation de la première liaison phosphodiester; l'élongation, pendant laquelle l'ARN polymérase synthétise la molécule d'ARN; et enfin la terminaison, au cours de laquelle l'ARN polymérase se décroche ou est éventuellement recyclée et l'ARN libéré. Dans la suite de ce chapitre, je présenterai les machineries de transcription des ARN polymérases II et III en me basant essentiellement sur les travaux réalisés chez la levure *S. cerevisiae*, et mettrai en avant certaines particularités propres aux eucaryotes pluricellulaires.

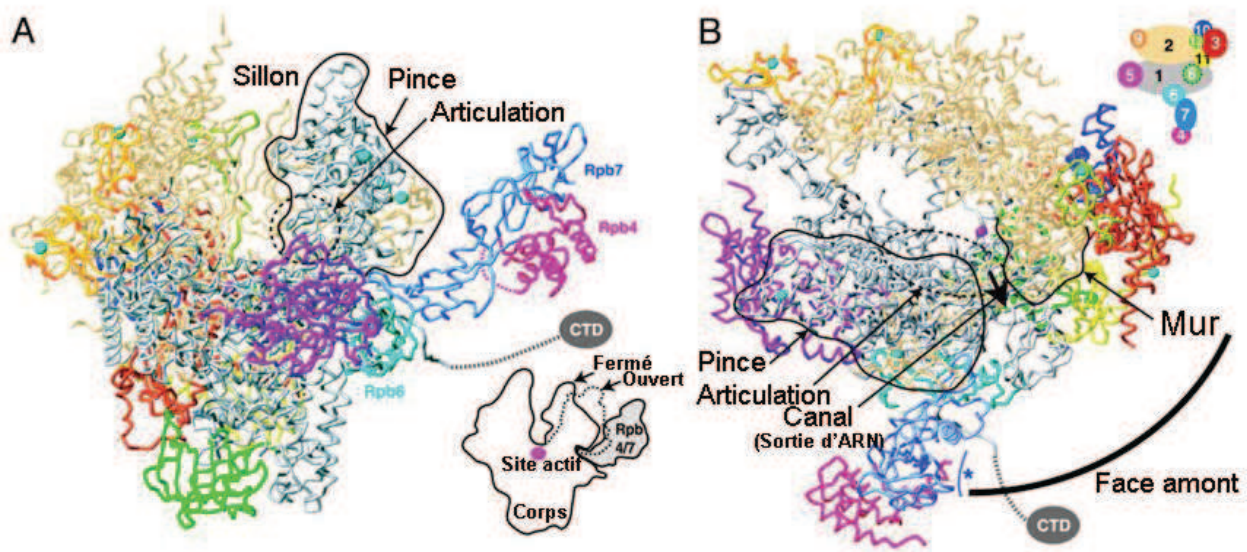
## a) Transcription par l'ARN polymérase II

### L'ARN polymérase II: structure et fonction

L'ARN polymérase II est une protéine hétéro-multimérique de 0,5 MDa composée de 12 sous-unités très bien conservées parmi l'ensemble archeo-eucaryotique, appelées Rpb1 à Rpb12 chez la levure *S. cerevisiae*. L'ARN polymérase II de *S. cerevisiae* peut être dissociée en un cœur catalytique de dix sous-unités et un hétérodimère composé des sous-unités Rpb4 et Rpb7 (Edwards et al., 1991). Cinq sous-unités (Rpb5, Rpb6, Rpb8, Rpb10, Rpb12) sont communes aux trois ARN polymérases, et sept sous-unités (Rpb1, Rpb2, Rpb3, Rpb4, Rpb7, Rpb9, Rpb11) ont des paralogues chez les ARN polymérases I et III. Les ARN polymérases I et III possèdent par ailleurs deux (Rpa34, Rpa49) et cinq (Rpc31, Rpc34, Rpc37, Rpc53, Rpc82) sous-unités spécifiques, respectivement. Dix des douze gènes codant les sous-unités de l'ARN polymérase II sont essentiels. Seuls *RPB4* et *RPB9* peuvent être délétés, bien que ces délétions provoquent des phénotypes de croissance cellulaire ralentie (Hampsey, 1998).

La structure cristalline de l'ARN polymérase II (Cramer et al., 2000, 2001) a été obtenue par l'équipe de R. Kornberg, qui a reçu le Prix Nobel de Chimie pour son travail sur les bases moléculaires de la transcription chez les eucaryotes. Cette structure, comportant les dix sous-unités du cœur de l'ARN polymérase II, a été obtenue à la fois seule, et avec l'ADN et l'ARN sous forme d'un complexe de transcription (Gnatt et al., 2001). Elle montre que les deux plus grandes sous-unités Rpb1 et Rpb2 constituent la masse centrale de l'enzyme et forment un sillon ("cleft") chargé positivement par lequel s'introduit l'ADN matrice (Figure 1). L'un des côtés du sillon formerait une pince mobile ("clamp") qui pourrait adopter une conformation ouverte ou fermée selon qu'elle interagit ou non avec l'ADN. Le site actif de l'ARN polymérase II est localisé au niveau du mur ("wall") définissant le fond du sillon. Un pore, situé sous le site actif, s'élargit vers la surface créant ainsi un entonnoir inversé ("funnel"). Le bord du pore est constitué d'une boucle de la sous-unité Rpb1 contenant un motif capable de lier un ion  $Mg^{2+}$  (métal A). Des travaux plus récents ont décrits la structure de l'ARN polymérase II complète, contenant les sous-unités Rpb4 et Rpb7 situées sous la pince mobile (Craighead et al., 2002; Armache et al., 2003, 2005; Bushnell & Kornberg, 2003). Le domaine C-terminal de la sous-unité Rpb1 de l'ARN polymérase II (CTD) n'est pas structuré, mais se situe probablement au niveau de la face amont, sous la pince mobile.

L'obtention de la structure de l'ARN polymérase II a permis de mieux comprendre son mode de fonctionnement. Deux études récentes se sont, par exemple, intéressées au mécanisme de sélection et d'incorporation des nucléosides tri-phosphates (NTP) dans le complexe d'élongation de l'ARN polymérase II (Westover et al., 2004; Kettenberger et al., 2004). Les auteurs ont proposé qu'il existe trois sites de liaison aux NTPs, localisés dans le pore à proximité de l'extrémité 3' de l'ARN néosynthétisé. Les NTPs pourraient entrer à travers le "funnel", et se lier à un premier site, dit d'entrée (site E pour "entry site"), puis atteindre le site de pré-insertion grâce à un mouvement de rotation. A cette étape, un mécanisme de contrôle prévient l'insertion de dNTPs ou de NTPs incorrects. Les NTPs passent alors dans le site d'insertion (site A pour "addition site") permettant la formation d'une liaison phosphodiester avec l'ARN néosynthétisé. Un modèle alternatif suggère que les NTPs pourraient se lier à la matrice d'ADN avant leur translocation dans le site actif, et atteindraient le site actif à travers le sillon (Gong et al., 2005).



**Figure 1. Structure de l'ARN polymérase II obtenue par diffraction des rayons X (d'après Armache et al., 2003).**

**A.** Vue de face de l'ARN polymérase II. Les sous-unités de l'ARN polymérase II sont représentées par des couleurs différentes (se référer au schéma situé en haut à droite: les sous-unités sont numérotées d'après la nomenclature usuelle de Rpb1 à Rpb12). **B.** Vue du dessus de l'ARN polymérase II. Les sous-unités de Pol II sont représentées par des couleurs différentes.

Entre les deux panneaux, un modèle hypothétique de changement de conformation de l'ARN polymérase II est présenté.



Des études de génétique et de biochimie ont pu éclairer la fonction de chacune des douze sous-unités de l'ARN polymérase II dans la transcription. La sous-unité Rpb9 appartient à une famille conservée de protéines se liant au zinc, incluant Rpc11 (ARN polymérase III), Rpa12 (ARN polymérase I) et dans une certaine mesure le facteur d'élongation TFIIS (voir aussi Chapitre II). Cette sous-unité joue un rôle à la fois dans l'initiation de la transcription par l'ARN polymérase II pour la sélection du site d'initiation de la transcription (Hull et al., 1995; Sun et al., 1996), et dans l'élongation de la transcription (Hemming et al., 2000), notamment en favorisant la fidélité de l'ARN polymérase II (Nesser et al., 2006).

De nombreuses mutations dans les sous-unités Rpb1 et Rpb2 ont été isolées et caractérisées. La plupart d'entre elles affectent l'initiation de la transcription par l'ARN polymérase II, et plus particulièrement le choix du site d'initiation (Hekmatpanah & Young, 1991; Berroteran et al., 1994). D'autres mutations confèrent une sensibilité à un analogue de nucléotide, le 6-azauracile (6AU), un phénotype associé à des défauts dans l'élongation de la transcription (Archambault et al., 1992; Powell & Reines, 1996). Cependant, dans le cas de la sous-unité Rpb1, ce phénotype pourrait être lié à la perte d'interaction de l'ARN polymérase II avec le facteur d'élongation TFIIS. En effet, la sensibilité au 6AU des mutants *rpb1* peut être supprimée par la surexpression de *DST1*, le gène codant la protéine TFIIS (Archambault et al., 1992) (voir aussi Chapitre II).

Une des caractéristiques de l'ARN polymérase II par rapport aux autres ARN polymérases, est la présence du domaine CTD de la sous-unité Rpb1, servant de plateforme de signalisation pour l'ARN polymérase II. Le CTD consiste, selon les espèces, en 25 à 52 répétitions de l'heptapeptide Tyr-Ser-Pro-Thr-Ser-Pro-Ser très bien conservé au cours de l'évolution. Cette séquence peut être modifiée post-traductionnellement de façon importante *in vivo*, notamment via des phosphorylations se produisant principalement sur les sérines 2 et 5 (Phatnani & Greenleaf, 2006; Chapman et al., 2008). Il est suggéré que les modifications du CTD peuvent servir à coordonner la transcription et la maturation de l'ARN pré-messager (Komarnitsky et al., 2000). Afin de former un complexe de préinitiation (PIC) compétent pour l'initiation au promoteur d'un gène donné, le CTD doit se trouver sous une forme hypophosphorylée. Après l'assemblage du PIC, les répétitions du CTD sont phosphorylées sur la sérine 5 par la kinase cycline-dépendant Kin28/Cdk7, une des sous-unités du facteur général de transcription TFIIF. Cet événement de phosphorylation provoque le recrutement de la machinerie de coiffage de l'ARN pré-messager au cours des étapes précoces de la transcription (Cho et al., 1997; Komarnitsky et al., 2000; Schroeder et al., 2000; Fabrega et

al., 2003). Au fur et à mesure que le complexe d'élongation de la transcription progresse, un changement apparaît dans les modifications du CTD, avec l'apparition de la phosphorylation de la sérine 2 sous l'action du complexe CTDK-I (Cho et al., 2001) et des protéines Bur1/Bur2 (Qiu et al., 2009) chez *S. cerevisiae* ou du facteur P-TEFb ("Positive Transcription-Elongation Factor-b") chez la plupart des autres eucaryotes (Wood & Shilatifard, 2006), et la disparition de la phosphorylation de la sérine 5 via différentes phosphatases (Meinhart et al., 2005; Mosley et al., 2009). Des expériences d'immunoprécipitation de la chromatine ont montré que l'augmentation de la phosphorylation de la sérine 2 a lieu au fur et à mesure que la transcription progresse le long de la phase ouverte de lecture (ORF) (Komarnitsky et al., 2000). Lorsque la machinerie de transcription approche de l'extrémité 3' de l'ORF, les complexes de terminaison et de polyadénylation sont recrutés en partie grâce à des interactions avec le CTD phosphorylé sur la sérine 2 (Licatalosi et al., 2002; Meinhart & Cramer, 2004; Kim et al., 2004). Il a récemment été mis en évidence que le CTD pouvait également être phosphorylé sur sa sérine 7. Cette phosphorylation serait importante pour la maturation de deux petits ARN nucléaires dans des lignées de cellules de mammifères (Chapman et al., 2007; Egloff et al., 2007), et serait également présente chez *S. cerevisiae* et effectuée par la sous-unité Kin28/Cdk7 de TFIIH (Akhtar et al., 2009).

### Le cycle de transcription par l'ARN polymérase II (Figure 2)

- la formation du complexe de préinitiation

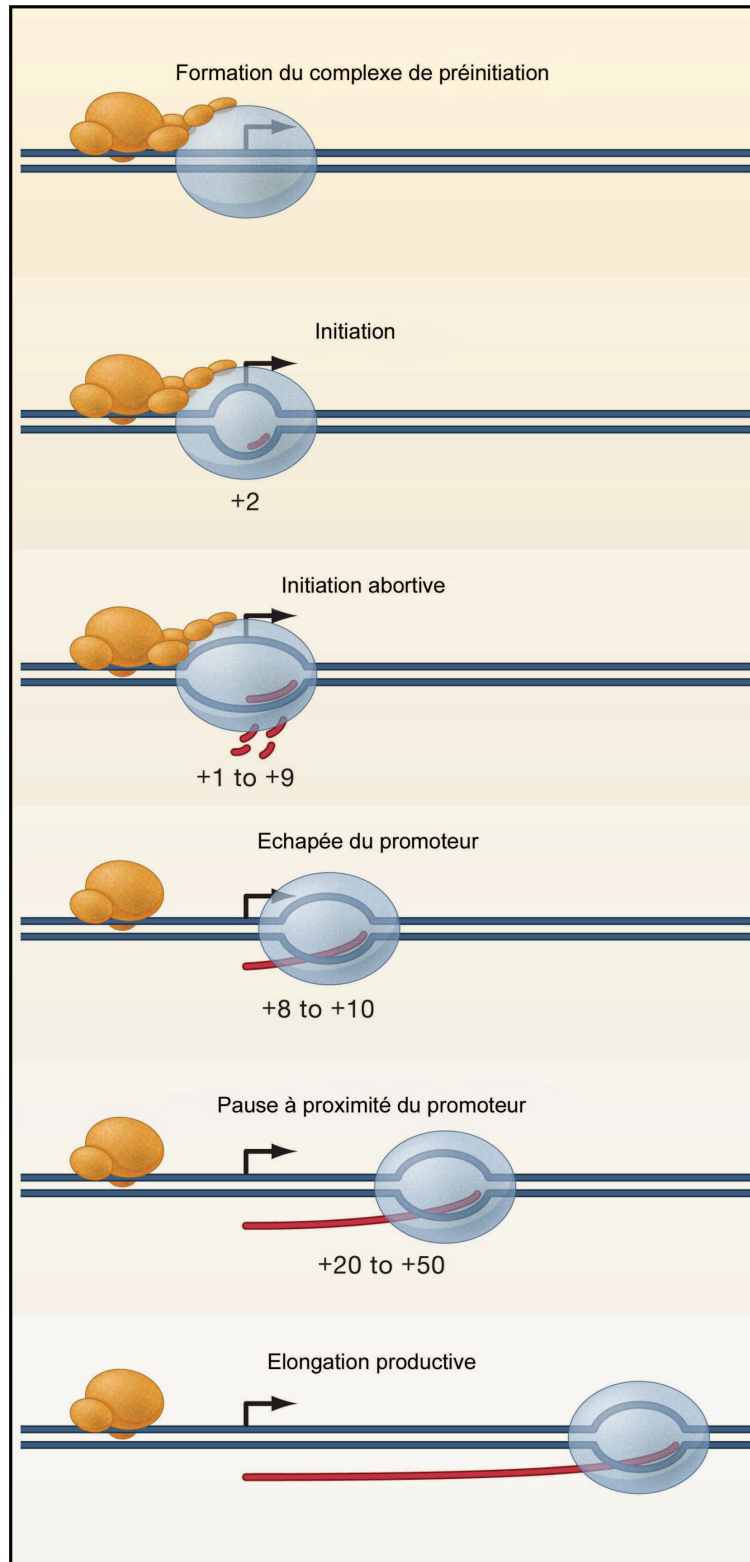
Des études *in vitro* ont montré que le complexe minimal requis pour l'initiation de la transcription, aussi appelé complexe de préinitiation (PIC), est composé de l'ARN polymérase II ainsi que des facteurs généraux de transcription TFIIA, TFIIB, TFIID, TFIIIE, TFIIF et TFIIH (Reese, 2003) (voir Tableau 1 pour le rôle des différents facteurs généraux). Les facteurs généraux de transcription interviennent dans la reconnaissance du promoteur, le recrutement de l'ARN polymérase II et l'ouverture de la bulle de transcription. Le promoteur peut être défini comme une séquence d'ADN située immédiatement en amont des gènes qui sert de plateforme d'assemblage au PIC. Chez les eucaryotes, il contient un ou plusieurs motifs caractéristiques, dont le plus connu est la boîte TATA. La protéine TBP ("TATA-Binding Protein"), une des sous-unités du complexe TFIID interagit avec la boîte TATA (Klein & Struhl, 1994). Il s'agit d'une séquence, située entre 25 et 30 paires de base en

amont du site de démarrage de la transcription chez la plupart des eucaryotes, excepté chez *S. cerevisiae* où la distance peut varier de 40 à 120 nucléotides. De nombreux promoteurs sont cependant dépourvus de boîte TATA. D'autres éléments (absents chez *S. cerevisiae*) tels que l'élément promoteur basal DPE (Downstream Promoter Element), l'élément de reconnaissance par TFIIB (BRE) ou l'élément initiateur (INR) (Juven-Gershon et al., 2008) ont été identifiés. Des études réalisées *in vitro* ont permis de proposer un ordre d'arrivée des facteurs généraux de la transcription, qui se fait de façon séquentielle (Buratowski et al., 1989; Hampsey, 1998). Le promoteur est reconnu par le complexe TFIID qui cause une courbure de l'ADN et agit comme une plateforme pour l'arrivée des autres facteurs généraux. TFIIA et TFIIB viennent ensuite stabiliser TFIID et recruter un complexe formé par TFIIF et l'ARN polymérase II. Enfin, TFIIIE puis TFIIH viennent s'associer à la structure préexistante, et l'activité hélicase de TFIIH catalyse l'ouverture de la bulle de transcription. Le complexe de transcription est alors appelé complexe ouvert. Une partie du travail effectué au cours de ma thèse a consisté à montrer que le Médiateur, en interagissant avec la sous-unité Rad3 du complexe TFIIH, est impliqué dans la formation du PIC *in vivo*, selon un schéma différent de celui décrit *in vitro*.

Ce PIC minimal n'est cependant pas suffisant pour répondre aux activateurs ou initier la transcription sur une matrice chromatinienne, mais nécessite l'intervention de différents coactivateurs tels que les complexes SAGA, Médiateur ou les complexes de remodelage de la chromatine (voir ci-dessous et Chapitre III).

**Tableau 1: Les facteurs généraux de la transcription.**

<i>Facteur</i>		<i>Nombre de sous-unités</i>	<i>Fonction</i>
TFIIA		2 / 3	Stabilise TBP et la liaison entre TFIID et l'ADN. Rôle de régulateur transcriptionnel.
TFIIB		1	Interagit avec la TBP, l'ARN polymérase II et la région promotrice. Joue un rôle dans le choix du site d'initiation.
TFIID	TBP	1	Interagit avec la boîte TATA et déforme la région promotrice. Plateforme pour l'assemblage de TFIIB, TFIIA et des TAFs.
	TAFs	14	Impliqué dans la reconnaissance des promoteurs dépourvus de boîte TATA.
TFIIF		3 / 2	Interagit avec l'ARN polymérase II et permet son recrutement dans le PIC. Rôle dans les étapes précoces de l'élongation.
TFIIIE		2	Interagit avec la région promotrice. Participe au recrutement de TFIIH.
TFIIH		10	Rôle dans la transcription et la réparation de l'ADN. Possède une activité kinase et deux activités hélicases.



**Figure 2. Les étapes de la transcription (d'après Margaritis & Holstege, 2008).**

Représentation schématique des premières étapes de la transcription. L'ARN polymérase II est représenté en bleu clair et les facteurs généraux de transcription en orange. Les flèches représentent le site de démarrage de la transcription. Les nombres se réfèrent à la position du site actif de l'ARN polymérase II sur la matrice d'ADN par rapport au site de démarrage de la transcription.

- l'échappée du promoteur ("promoter escape")

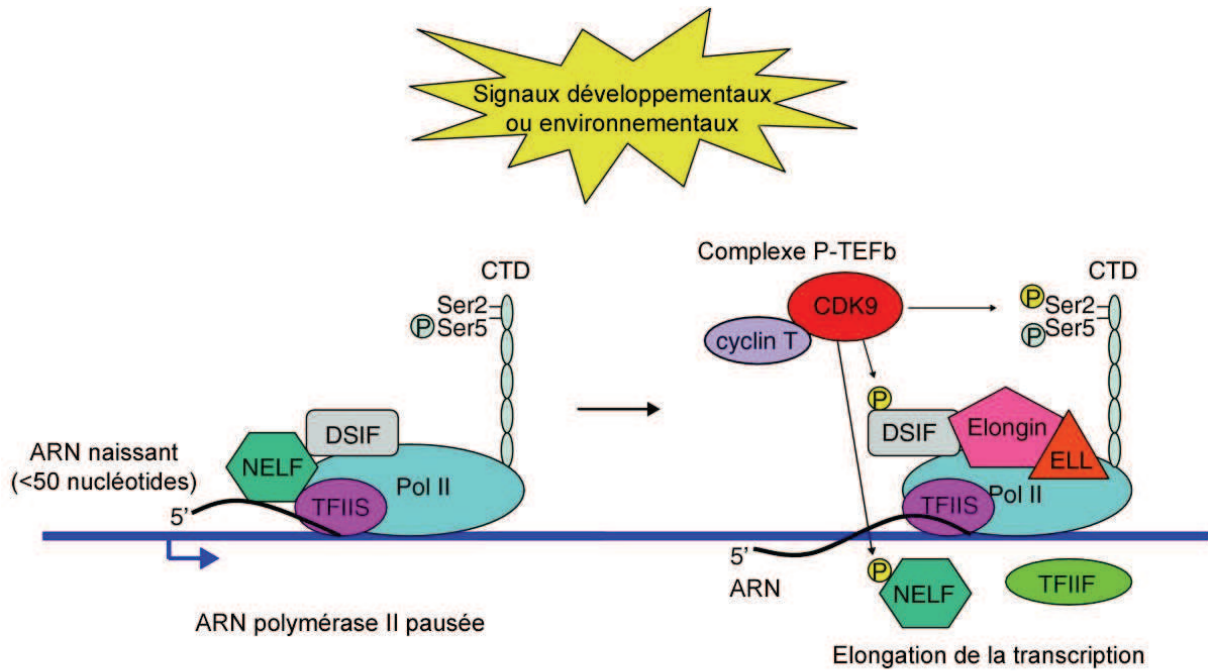
L'échappée du promoteur décrit une étape précoce de l'élongation au cours de laquelle l'ARN polymérase II rompt ses contacts avec les séquences promotrices et une partie des facteurs liés au promoteur (Saunders et al., 2006; Wade & Struhl, 2008; Margaritis & Holstege, 2008). *In vitro*, cette étape peut être limitante pour la transcription à partir de certains promoteurs. Il a été montré qu'*in vitro*, l'échappée du promoteur est généralement précédée de plusieurs cycles d'initiation abortive, pendant laquelle l'ARN polymérase II synthétise de courts ARNs de quelques nucléotides (Lescure et al., 1981; Hsu, 2009).

- la pause à proximité du promoteur ("promoter proximal pausing")

La pause à proximité du promoteur est un phénomène au cours duquel l'ARN polymérase II est en situation de pause dans la région 5' de l'unité de transcription, et ne peut progresser efficacement vers une élongation productive qu'en réponse à une stimulation par des signaux extérieurs. Cette étape est un point important de régulation de l'élongation de la transcription par l'ARN polymérase II *in vivo* (voir aussi Chapitre III). La pause à proximité du promoteur est un phénomène très répandu qui régule la transcription de nombreux gènes chez les eucaryotes. Il a été particulièrement bien étudié chez *Drosophila melanogaster* dans le cas des gènes dont l'expression est induite par un choc thermique (Saunders et al., 2006; Gilmour, 2009). Ce type de régulation a également été mis en évidence pour les proto-oncogènes *MYC* et *FOS* chez les mammifères (Kim et al., 2005). La pause pourrait à la fois servir de point de contrôle pour la vérification de la préparation de l'ARN polymérase II à son entrée en élongation productive, et permettre une régulation rapide de l'expression des gènes. Par ailleurs, cette étape pourrait faciliter l'ajout de la coiffe sur les ARN néosynthétisés. Les mécanismes de régulation de cette pause sont encore mal connus, cependant il apparaît que les complexes DSIF ("DRB Sensitivity-Inducing Factor") et NELF ("Negative Elongation Factor"), des répresseurs de l'élongation de la transcription, puisse être impliqués (Saunders et al., 2006).

- l'élongation productive

L'entrée en élongation productive coïncide avec le recrutement du complexe P-TEFb, dont la kinase Cdk9 est responsable de la phosphorylation de DSIF (homologue des protéines Spt4 et Spt5 chez *S. cerevisiae*), NELF (absent chez *S. cerevisiae*) et de la sérine 2 du CTD de l'ARN polymérase II (Saunders et al., 2006) (Figure 3). NELF quitte alors le complexe d'élongation, alors que DSIF y reste associé et joue à présent un rôle positif dans l'élongation



**Figure 3. ARN polymérase II en situation de pause au promoteur et transition vers l'étape d'élongation (d'après Wu & Snyder, 2008).**

L'ARN polymérase II (Pol II) peut être en situation de pause à proximité du promoteur de nombreux gènes, après avoir entamé une initiation de la transcription produisant un court ARN. L'échappée du promoteur peut être induite par des signaux développementaux ou environnementaux. Dans le complexe pausé, seule la sérine 5 du CTD est phosphorylée. Le complexe P-TEFb (composé de Cdk9 et de la cycline T) facilite le passage de l'ARN polymérase II en élongation productive en phosphorylant DSIF, NELF et la sérine 2 du CTD.

de la transcription (Andrulis et al., 2000). D'autres facteurs agissant sur la vitesse d'élongation du complexe de transcription peuvent entrer en jeu. Les protéines TFIIF, ELL ("eleven-nineteen lysine-rich in leukemia") et les elongines stimulent l'élongation de la transcription *in vitro* lors des pauses transitoires que peut rencontrer l'ARN polymérase II. TFIIF est vraisemblablement plutôt impliqué dans les étapes précoces de l'élongation (Shilatifard et al., 2003). Par ailleurs, la protéine TFIIS aide l'ARN polymérase II à traverser des sites de blocage de la transcription (voir aussi Chapitre II). D'autres facteurs tels que le complexe THO/TREX facilitent l'élongation de la transcription en empêchant des appariements inappropriés entre la matrice d'ADN et l'ARN néosynthétisé (Huertas & Aguilera, 2003; Rondón et al., 2003). Enfin, d'autres facteurs comme les complexes FACT, Elongator ou la protéine Spt6, et dans une moindre mesure les complexes de remodelage de la chromatine,

facilitent le passage de l'ARN polymérase II à travers la chromatine en désassemblant et réassemblant les nucléosomes le long de l'unité de transcription (Saunders et al., 2006). Le complexe d'élongation de l'ARN polymérase II fonctionne, via le CTD phosphorylé sur la sérine 2, comme une plateforme coordonnant la transcription, la maturation des ARN messagers et leur export.

- la terminaison de la transcription

La terminaison de la transcription par l'ARN polymérase II est une étape essentielle permettant d'éviter les interférences avec les gènes en amont, dont le mécanisme est relié à celui de la maturation de l'extrémité 3' des transcrits. Deux modèles ont été proposés pour expliquer ce lien. Le modèle "anti-terminateur" suggère que les séquences de poly-adénylations de l'ARN entraînent une modification des facteurs associés à l'ARN polymérase, la rendant moins processive, conduisant ainsi à l'arrêt de la transcription. Un second modèle appelé "torpedo" propose que le clivage du transcrit par la machinerie de poly-adénylation génère une nouvelle extrémité 5' qui serait le substrat d'une nucléase dont l'activité entraîne la dissociation de l'ARN polymérase II (Buratowski, 2005). Par ailleurs, un mode de terminaison alternatif a récemment été mis en évidence dans le cas des ARN non-codants. Celui-ci fait intervenir le complexe Ndr1, qui reconnaît des sites spécifiques à l'extrémité 3' de l'ARN. Ces ARNs sont généralement plus ou moins rapidement dégradés par l'exosome et TRAMP (Lykke-Andersen & Jensen, 2007).

- la réinitiation (Figure 4)

Après l'initiation de la transcription par l'ARN polymérase II *in vitro*, un certain nombre de facteurs généraux de la transcription peuvent rester au niveau du promoteur dans un complexe appelé le "scaffold" (plateforme de réinitiation, se traduit littéralement par échafaudage). Ce complexe, uniquement identifié *in vitro*, serait formé du Médiateur et des facteurs généraux de la transcription à l'exception de TFIIB et TFIIF. Ce complexe existe probablement dans le cas des gènes les plus transcrits et permettrait d'éviter l'étape relativement lente du recrutement des facteurs généraux de transcription pour les cycles de transcription ultérieurs. Le complexe "scaffold" pourrait alors recruter rapidement les facteurs généraux manquants et permettre la réinitiation de la transcription (Yudkovsky et al., 2000; Hahn, 2004).

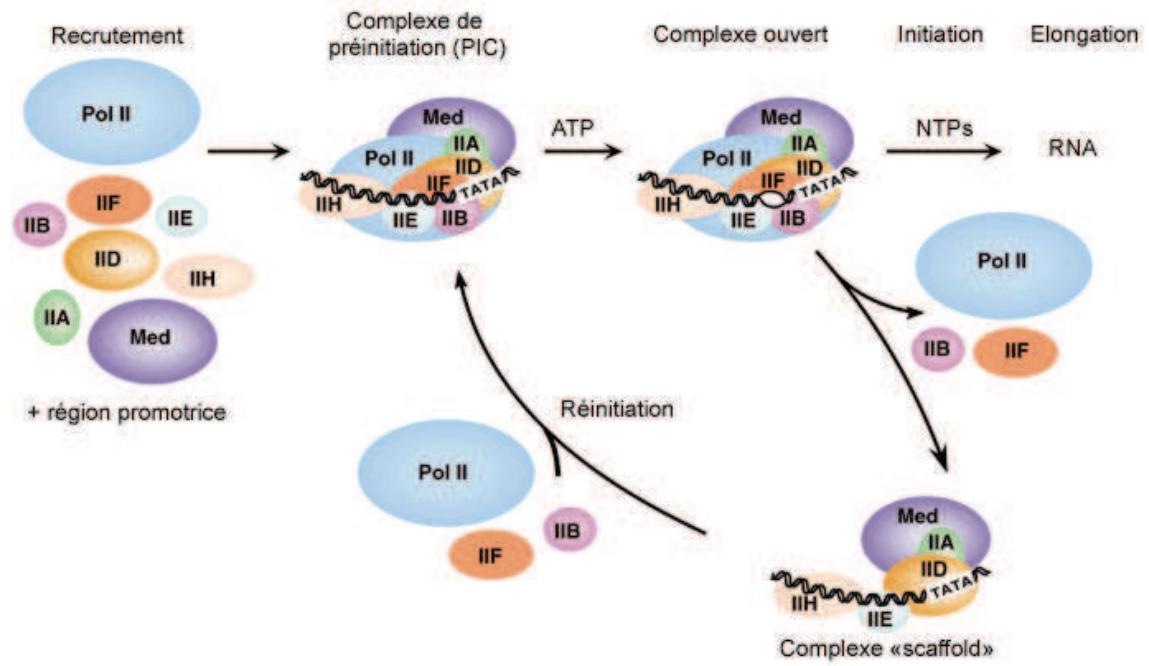
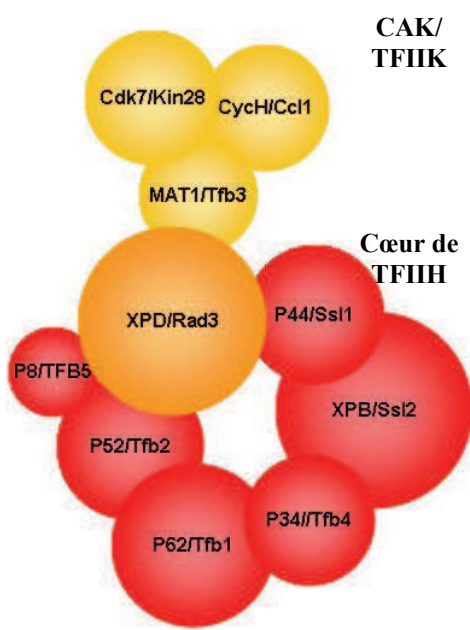


Figure 4. L'initiation et la réinitiation de la transcription par l'ARN polymérase II (d'après Hahn, 2004).



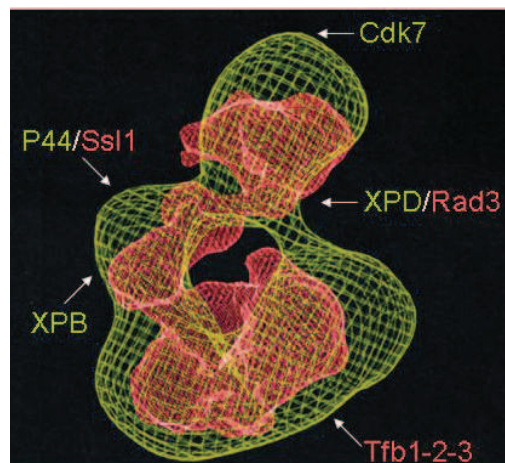
Le facteur général de transcription TFIID**Tableau 2: Le complexe TFIID chez l'homme et la levure *S. cerevisiae*.**


	Homme	Levure	Poids moléculaire Hs (kDa)	Fonctions
<b>CAK/TFIIK</b>	Cdk7	Kin28	41	Kinase de la transcription et du cycle cellulaire Cycline, régule l'activité de la kinase Cdk7/Kin28 Régule la spécificité de substrat de la kinase Cdk7/Kin28
	CycH	Ccl1	38	
	MAT1	Tfb3	32	
<b>Cœur de TFIID</b>	XPB	Ssl2	89	Hélicase 3'-5' ATP dépendante Hélicase 5'-3' ATP dépendante. Ancre le sous-complexe CAK/TFIIK au cœur de TFIID
	XPB	Rad3	80	
	p62	Tfb1	62	- Ancre XPB au cœur de TFIID
	p52	Tfb2	52	
	p44	Ssl1	44	Active XPD. Rôle dans l'échappée du promoteur par l'ARN polymérase II
	p34	Tfb4	34	
	p8	Tfb5	8	- Stabilise le complexe. Stimule l'activité ATPase de XPB

TFIID est un complexe multi-protéique très conservé au cours de l'évolution, composé de 10 sous-unités détaillées dans le Tableau 2 (Ranish et al., 2004; Thomas & Chiang, 2006). Il possède trois activités enzymatiques requises lors de la transcription: une activité ATPase ADN-dépendante (Conaway & Conaway, 1989; Roy et al., 1994), une activité hélicase ATP-dépendante (Schaeffer et al., 1993) et une activité CTD kinase (Feaver et al., 1991; Serizawa et al., 1992; Lu et al., 1992). Les activités ATPase et hélicase de TFIID sont requises pour l'initiation de la transcription et l'échappée du promoteur. Elles permettent la formation du complexe de transcription dit "ouvert" et celle des premières liaisons phosphodiester de la molécule d'ARN. L'activité kinase de TFIID est responsable de la phosphorylation de la sérine 5 du CTD de l'ARN polymérase II, étape essentielle à la transition entre l'initiation et l'élongation de la transcription. En plus de son rôle dans la transcription, TFIID est impliqué dans la réparation par excision de nucléotides (NER, "nucleotide excision repair") (Prakash & Prakash, 2000) et le contrôle du cycle cellulaire (Fisher, 2005), bien que cette dernière activité ne semble pas être dépendante de TFIID chez *S. cerevisiae*. TFIID peut être séparé en deux sous-complexes: CAK/TFIIK et le cœur de

TFIIH, lesquels sont reliés par l'hélicase XPD/Rad3. Chez l'homme, des mutations dans les sous-unités XPD/Rad3 ou XPB/Ssl2 causent de graves maladies génétiques telles que le xeroderma pigmentosum, la trichotiodystrophie ou le syndrome de Cockayne (Lehmann, 2001).

L'obtention de modèles structuraux des complexes TFIIH chez l'homme et la levure *S. cerevisiae* par microscopie électronique à une résolution comprise entre 15 et 40 Å, a mis en évidence la grande conservation de l'organisation de TFIIH chez les eucaryotes (Chang & Kornberg, 2000; Schultz et al., 2000). TFIIH apparaît sous la forme d'un anneau surmonté par une protubérance. L'anneau est constitué du cœur de TFIIH. On y retrouve les deux hélicases XPB/Ssl2 et XPD/Rad3 de part et d'autre du module. Les quatre autres protéines du cœur de TFIIH complètent la structure. La protubérance dans laquelle est localisée la protéine Cdk7, correspond à CAK/TFIIK qui se greffe sur le cœur *via* XPD/Rad3. La superposition des modèles de TFIIH humain et de levure montre que les complexes adoptent la même organisation et témoigne ainsi d'une similitude remarquable (Figure 5).



**Figure 5. Modèle de microscopie électronique des complexes TFIIH humain et de levure (d'après Chang & Kornberg, 2000; Schultz et al., 2000).**

En jaune: Modèle de TFIIH humain à 38 Å de résolution. En rouge: Modèle de TFIIH de levure à 18 Å de résolution. Les positions des protéines obtenues par un immuno-localisation sont indiquées en jaune pour les protéines humaines et en rouge pour les protéines de levure.

## La régulation de la transcription par l'ARN polymérase II

Il est aujourd'hui bien connu que différents facteurs de transcription sont capables d'activer ou de réprimer la transcription en se fixant à des éléments régulateurs plus ou moins proches du promoteur. Les activateurs et répresseurs de la transcription sont généralement des protéines modulaires composées d'un domaine de liaison à l'ADN et d'un domaine d'activation (ou de répression) (Triezenberg, 1995). Les signaux d'activation ou de répression peuvent être transmis à la machinerie basale de transcription via des coactivateurs.

- les coactivateurs

Les coactivateurs sont de larges complexes multiprotéiques capables de servir d'intermédiaire entre les activateurs spécifiques de la transcription et la machinerie basale de transcription ou/et de permettre à la machinerie de transcription de franchir la barrière nucléosomale grâce à une activité de remodelage ou de modification de la chromatine (Näär et al., 2001). Une première classe de coactivateurs est composée de complexes agissant sur la machinerie basale de transcription. Parmi ceux-ci, on retrouve le complexe TFIID et le Médiateur de la transcription (voir ci-dessous). En plus de son rôle de facteur général de la transcription, TFIID serait impliqué, via les TAFs ("TBP Associated Factors"), dans la réponse à différentes classes d'activateurs (Chen et al., 1994). Cependant, certains TAFs ne sont pas requis *in vivo* pour la transcription de l'ensemble des gènes (Näär et al., 2001; Thomas & Chiang, 2006). Les complexes de remodelage de la chromatine et de modification des histones forment une deuxième grande catégorie de coactivateurs (voir aussi Chapitre III). Les complexes de remodelage de la chromatine peuvent être classés en quatre principales familles, SWI/SNF, ISWI, CHD et INO80, capables d'hydrolyser l'ATP pour altérer les contacts entre les histones et l'ADN (Clapier & Cairns, 2009). Cette activité permettrait d'altérer les structures de la chromatine, afin de favoriser la liaison des activateurs et de la machinerie basale de transcription au niveau des régions promotrices. Un grand nombre de complexes de modification des histones contiennent des TAFs. C'est par exemple le cas du complexe SAGA ("Spt-Ada-Gnc5 Acetyltransferase") chez *S. cerevisiae*, qui possède une activité histone acétyltransférase via la protéine Gcn5 (Grant et al., 1997).

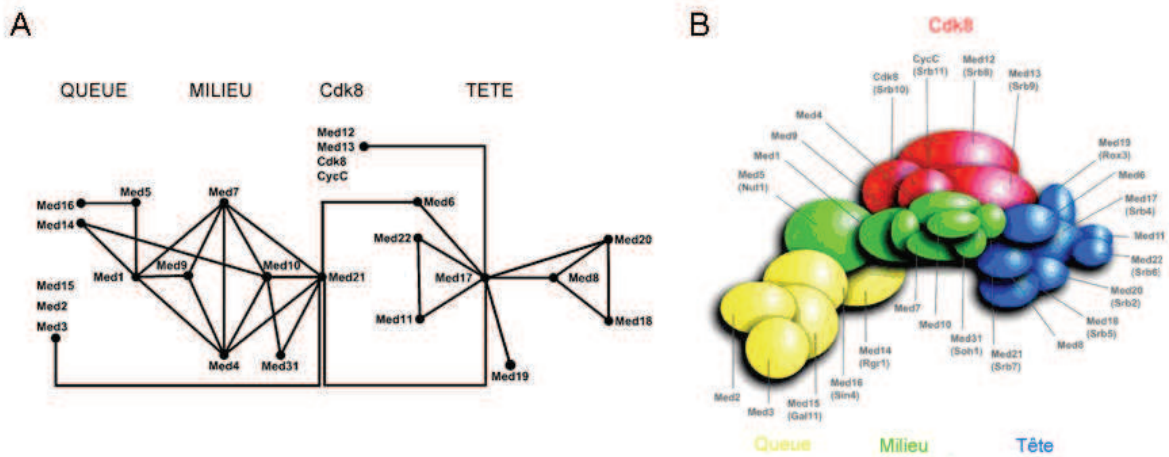
- le Médiateur de la transcription

Le Médiateur est un complexe protéique de très grande taille, composant essentiel de la machinerie de transcription par l'ARN polymérase II. Il joue un rôle clé à la fois en

stimulant la transcription basale, mais aussi en régulant la synthèse d'ARN messager, puisqu'il transmet les signaux de régulation vers l'ARN polymérase II. De ce fait, le Médiateur peut être considéré à la fois comme un facteur général de transcription et comme un coactivateur. Le Médiateur a été initialement purifié chez *S. cerevisiae* comme un complexe capable de stimuler, *in vitro*, la transcription d'un gène en présence d'un activateur (Kelleher et al., 1990; Flanagan et al., 1991; Kornberg, 2005). Chez *S. cerevisiae*, il est composé de 25 sous-unités, dont 10 sont essentielles (Tableau 3). Certaines sous-unités du Médiateur, les protéines Srb (Suppressor of RNA polymerase B), avaient été précédemment identifiées par des études de génétique comme des suppresseurs extragéniques d'une troncature du CTD de l'ARN polymérase II (Nonet & Young, 1989). En effet, l'un des rôles essentiels du Médiateur est la stimulation de la phosphorylation du CTD par TFIIH, étape clé de la transition entre l'initiation et l'élongation de la transcription (Kim et al., 1994; Max et al., 2007).

Une carte des interactions entre les sous-unités du Médiateur a été obtenue par un crible double-hybride chez *S. cerevisiae*, permettant ainsi de proposer un modèle pour l'organisation du complexe (Guglielmi et al., 2004) (Figure 6). Les sous-unités du Médiateur peuvent être séparées en quatre modules appelés module de tête, milieu, queue et module Cdk8. Le rôle du module Cdk8 est encore mal compris, mais il aurait principalement un effet répresseur sur la transcription par l'ARN polymérase II via son activité kinase phosphorylant le Médiateur lui-même, le complexe TFIIH et le CTD de l'ARN polymérase II (Hengartner et al., 1998; Akoulitchiev et al., 2000; Hallberg et al., 2004). Par ailleurs, ce module pourrait ne pas être systématiquement associé avec le reste du complexe (Björklund & Gustafsson, 2005; Andrau et al., 2006).

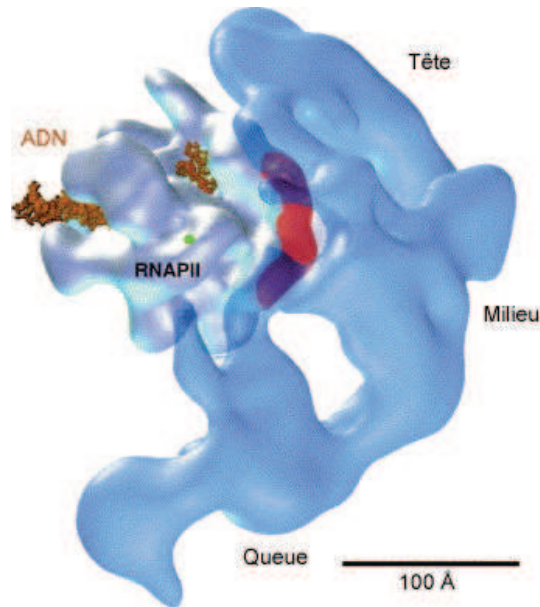
Les techniques actuelles ne permettent pas l'obtention d'un cristal de l'ensemble du Médiateur. Néanmoins, il est possible de résoudre à l'échelle atomique, la structure de sous-complexes du Médiateur impliquant une, deux ou trois protéines. Quatre modèles ont ainsi été obtenus à partir des cristaux de l'hétérodimère Med7-Med21 (Baumli et al., 2005), des sous-unités Med8-Med18-Med20 associées à TBP (Lariviere et al., 2006), du sous-module formé par la partie N-terminale de Med7 et Med31 (Koschubs et al., 2009), et de la cycline CycC, dont un modèle d'interaction avec la sous-unité Cdk8 a été obtenu par homologie (Hoeppner et al., 2005). Un modèle de l'association de Médiateur à l'ARN polymérase II a été obtenu par microscopie électronique à une résolution de 40 Å (Asturias et al., 1999; Davis et al., 2002). Celui-ci suggère que le Médiateur englobe l'ARN polymérase II et interagit avec elle via ses modules de tête et du milieu (Figure 7). Un changement de conformation a été mis en



**Figure 6. Organisation du Médiateur (d'après Guglielmi et al., 2004).**

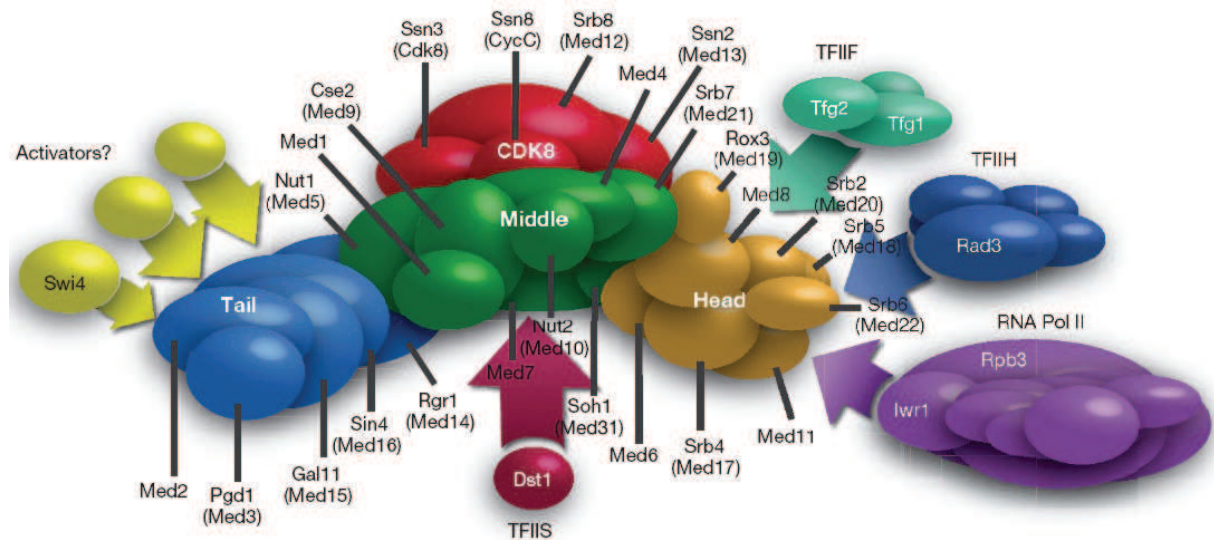
**A.** Carte des interactions entre les sous-unités du Médiateur. **B.** Schéma de l'organisation du complexe Médiateur défini par double-hybride.

évidence entre la forme libre du Médiateur, dite fermée, présentant un repliement de la queue sur le milieu, et une forme ouverte, lorsque le Médiateur est liée à l'ARN polymérase II. Ce changement de conformation serait important pour l'activation du complexe. Le Médiateur est également impliqué dans de nombreuses interactions protéine-protéine, notamment avec les autres composants du complexe de préinitiation. Ainsi, des interactions génétiques entre le module du milieu et le facteur TFIIS, ou le module de tête et les facteurs TFIIF et TFIIH ont été identifiées (Collins et al., 2007). De plus, le module de queue du Médiateur serait plus particulièrement impliqué dans la réponse aux activateurs spécifiques (Myers et al., 1999; Lee et al., 1999) (Figure 8).



**Figure 7. Interaction du Médiateur avec l'ARN polymérase II (d'après Chadick & Asturias, 2005).**

De nombreux contacts entre le Médiateur et l'ARN polymérase II (RNAPII) peuvent être établis au sein du complexe, impliquant principalement les modules de tête et du milieu du Médiateur, autour des sous-unités Rpb3-Rpb11 (en rouge) de l'ARN polymérase II. Le point vert indique la position du début du CTD de la sous-unité Rpb1 de l'ARN polymérase II.



**Figure 8. Carte d'interactions génétiques du Médiateur chez *S. cerevisiae* (d'après Collins et al., 2007).**

Carte des interactions fonctionnelles entre les modules du Médiateur et les intervenants de la transcription établie à partir d'une analyse bioinformatique qui regroupe les profils d'interactions génétiques réalisés avec 743 gènes chez *S. cerevisiae*.

Une étude de génomique fonctionnelle réalisée chez *S. cerevisiae* a révélé que, comme l'ARN polymérase II, le Médiateur est essentiel à la transcription de l'ensemble du génome *in vivo* (Holstege et al., 1998). En effet, l'analyse d'un mutant thermosensible de la sous-unité Med17 du module de tête, a montré une chute de l'expression de 93% du génome, ce qui est comparable à un mutant de l'ARN polymérase II. Des mutations affectant d'autres sous-unités du Médiateur ont des effets moins forts, alors que des mutants du module Cdk augmentent la transcription d'un certain nombre de gènes. Ce résultat témoigne de l'importance du Médiateur dans la transcription basale et activée, du rôle de régulateur négatif du module Cdk8, et suggère que certaines sous-unités ont une activité différente selon les gènes. Ceci a été confirmé par l'étude de l'effet transcriptionnel de la délétion des quinze sous-unités non-essentiels du Médiateur chez *S. cerevisiae* (van de Peppel et al., 2005), qui peuvent être classées en trois catégories: les sous-unités ayant un rôle positif sur la transcription, celles ayant un rôle négatif et celles dont la délétion n'affecte pas la transcription dans les conditions testées. Par ailleurs, il a été mis en évidence que certaines sous-unités peuvent avoir un effet antagoniste. C'est par exemple le cas entre le module Cdk et certains composants de la queue (Med15, Med2 et Med3), la tête (Med20 et Med18) ou le milieu (Med31), indiquant que différentes sous-unités d'un même complexe peuvent avoir des rôles opposés sur un même gène.

Des expériences de localisation à grande échelle réalisées chez *S. cerevisiae* et *S. pombe* ont révélé la présence de tous les modules du Médiateur en amont de l'ensemble des gènes activement transcrits, mais également de certains gènes non transcrits (Andrau et al., 2006; Zhu et al., 2006) (voir aussi Chapitre III). Cependant ce résultat pourrait dépendre des conditions de croissance, suggérant que la fonction du Médiateur pourrait être en partie promoteur-spécifique (Fan et al., 2006). Par ailleurs, il a été suggéré que le Médiateur pourrait également être présent au niveau de la région codante de certains gènes, bien que le rôle fonctionnel de cette occupation ne soit pas connu et que ce résultat n'ait pu être confirmé par la suite (Fan & Struhl, 2009).

Le Médiateur est un complexe conservé chez tous les eucaryotes, bien que le degré de conservation varie en fonction des sous-unités (Tableau 3). Les sous-unités les plus conservées forment le cœur du Médiateur. La composition sous-unitaire du Médiateur peut varier selon les espèces (Conaway et al., 2005). Il existe, par exemple, chez l'homme plusieurs complexes Médiateurs, dont les différences pourraient être liées à la régulation tissu-spécifique de l'expression des gènes. Une nomenclature unifiée de tous les complexes Médiateurs a récemment été adoptée, simplifiant ainsi les comparaisons inter-espèces



(Bourbon et al., 2004; Bourbon, 2008) (Tableau 3). Les structures du Médiateur humain et de levure sont très similaires, et les réarrangements structuraux du Médiateur de *S. cerevisiae* permettant son interaction avec l'ARN polymérase II sont tout à fait comparables aux réarrangements du Médiateur humain, suggérant que la dynamique structurale du Médiateur constitue un élément important de son mécanisme de régulation (Cai et al., 2009).

**Tableau 3: Sous-unités du Médiateur chez *S. cerevisiae* (d'après Bourbon, 2008).**

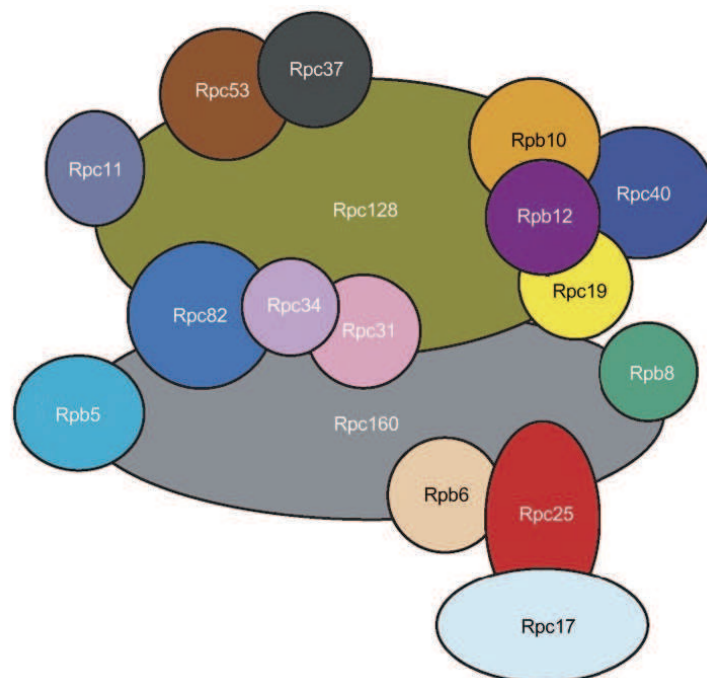
Dans la colonne de gauche, les sous-unités les plus conservées sont représentées en rouge, celles moyennement conservées en jaune orangé, celles non conservées en noir. Les sous-unités essentielles sont soulignées.

Nom systématique	<i>S. cerevisiae</i>	Module
<u>MED1</u>	Med1	Milieu
<u>MED2/29</u>	Med2	Queue
<u>MED3/27</u>	Pgd1	Queue
<u>MED4</u>	Med4	Milieu
MED5/24	Nut1	Queue
<u>MED6</u>	Med6	Tête
<u>MED7</u>	Med7	Milieu
<u>MED8</u>	Med8	Tête
<u>MED9</u>	Cse2	Milieu
<u>MED10</u>	Nut2	Milieu
<u>MED11</u>	Med11	Tête
<u>MED12</u>	Srb8	CDK
<u>MED13</u>	Srb9	CDK
<u>MED14</u>	Rgr1	Queue
<u>MED15</u>	Gal11	Queue
<u>MED16</u>	Sin4	Queue
<u>MED17</u>	Srb4	Tête
<u>MED18</u>	Srb5	Tête
<u>MED19</u>	Rox3	Tête
<u>MED20</u>	Srb2	Tête
<u>MED21</u>	Srb7	Milieu
<u>MED22</u>	Srb6	Tête
MED23	Absente	Queue
MED25	Absente	Non assigné
MED26	Absente	Non assigné
MED28	Absente	Non assigné
MED30	Absente	Non assigné
<u>MED31</u>	Soh1	Milieu
<u>Cdk8</u>	Srb10	CDK
<u>CycC</u>	Srb11	CDK



## b) Transcription par l'ARN polymérase III

L'ARN polymérase III est composée de 17 sous-unités dont 5 lui sont spécifiques (Rpc31, Rpc34, Rpc37, Rpc53, Rpc82) (Figure 9). Sa structure a été obtenue par microscopie électronique chez *S. cerevisiae* à une résolution de 17 Å et comparée à celle de l'ARN polymérase II (Fernández-Tornero et al., 2007). Il apparaît que l'ARN polymérase III est caractérisée par une tige formée par les sous-unités Rpc17/25 ("stalk", structure équivalente à celle formée par les sous-unités Rpb4/7 de l'ARN polymérase II) plus volumineuse ainsi que des protubérances au niveau du sillon ("cleft"), qui pourraient être attribuées aux sous-unités spécifiques à cette enzyme. La structure cristallographique du sous-complexe formé par les sous-unités Rpc17/25 a été obtenus (Jasiak et al., 2006). Les auteurs ont proposés un modèle de la structure de 11 sous-unités de l'ARN polymérase III par homologie avec celle de l'ARN polymérase II, révélant entre autres des particularités dans la région responsable de l'assemblage du complexe d'initiation. Des expériences de spectroscopie de masse ont par ailleurs confirmé le modèle d'organisation de l'ARN polymérase III (Werner et al., 1992, 1993; Flores et al., 1999), qui est composée d'un cœur de 10 sous-unités homologues à l'ARN polymérase II et des hétérodimères périphériques Rpc17/25, Rpc53/37 et Rpc82/34 (Lorenzen et al., 2007). La sous-unité Rpc31 permet de relier les dimères Rpc82/34 et Rpc17/25 au cœur de l'ARN polymérase III.



**Figure 9. Modèle d'organisation spatiale de l'ARN polymérase III de *S. cerevisiae* (d'après Flores et al., 1999; Lorenzen et al., 2007).**

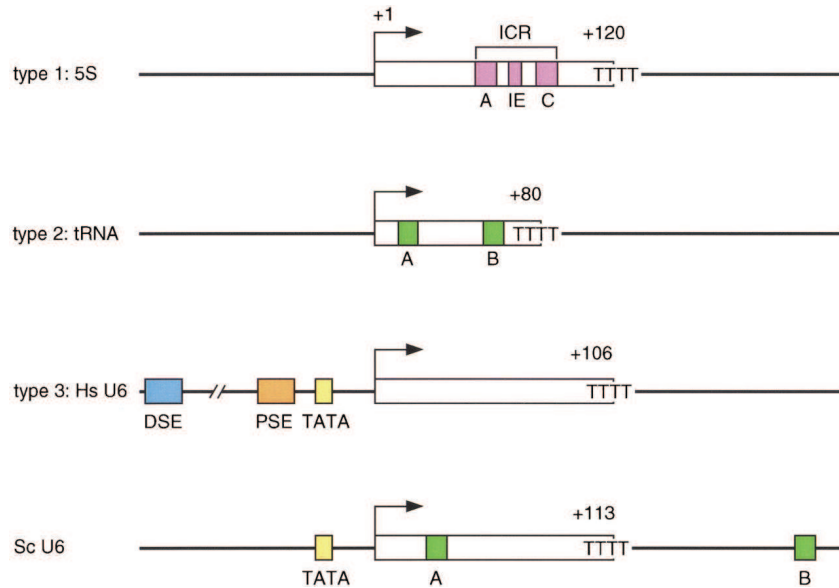
En plus des ARN de transfert (ARNt) et de l'ARN ribosomal 5S, l'ARN polymérase III assure la transcription de plusieurs ARNs non-codants, ayant tous pour caractéristique leur taille particulièrement courte (inférieure à 400 pb). Leur liste exacte a été établie chez la levure *S. cerevisiae* (Harismendy et al., 2003; Roberts et al., 2003; Moqtaderi & Struhl, 2004) (Tableau 4) (voir aussi Chapitre III). En plus des gènes transcrits par l'ARN polymérase III, huit nouveaux loci de rôle inconnu appelées *ETC 1* à *8* (pour "extra TFIIC") ont été identifiées comme étant occupées par le facteur TFIIC mais pas par l'ARN polymérase III ou ses autres facteurs (Moqtaderi & Struhl, 2004). Cependant *ETC5*, appelé également *RNA170* semble être transcrit et occupé faiblement par l'ARN polymérase III (Olivas et al., 1997; Harismendy et al., 2003). Chez les mammifères, la liste précise des gènes transcrits par l'ARN polymérase III n'est pas connue. Cependant nous pouvons citer l'ARN de la RNase MRP, l'ARN 7SK ou les SINE ("Short Interspersed repeated DNA elements") transcrits par l'ARN polymérase III mais absents chez *S. cerevisiae* (Dieci et al., 2007).

**Tableau 4: Transcrits de l'ARN polymérase III chez *S. cerevisiae*.**

Nom	Rôle
ARNt	Traduction des ARNm. 42 familles d'isoaccepteurs chez <i>S. cerevisiae</i>
ARNr 5S	S'associe avec la grande sous-unité du ribosome
snARN U6 (SNR6)	Petit ARN nucléaire impliqué dans l'épissage
ARN 7SL (SCR1)	Impliqué dans la formation de la "signal recognition particule"
ARN de RNase P (RPR1)	Ribonucléase clivant l'extrémité 5' des pré-ARNt (H1 chez la souris)
SNR52	Petit ARN nucléolaire
RNA170 (ETC5)	Inconnu.
ZOD1	Inconnu. N'est pas transcrit en conditions standard.

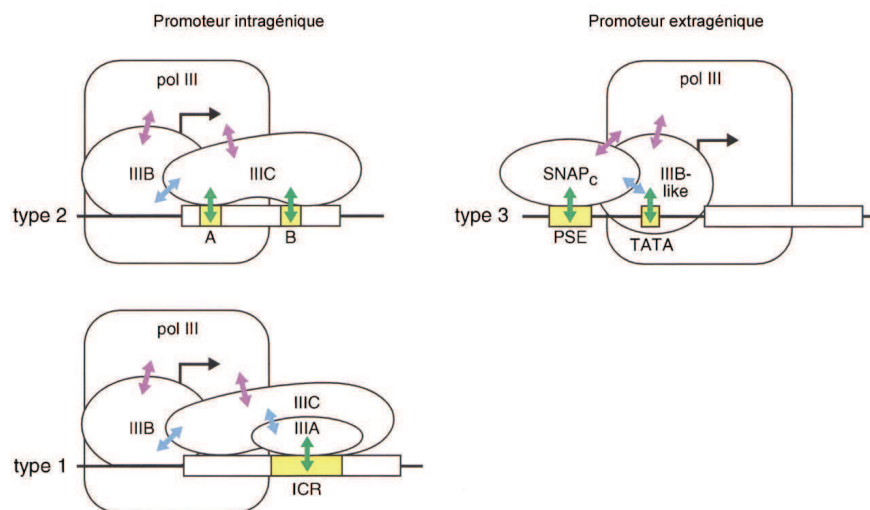
Il existe 3 types de promoteurs pour la transcription par l'ARN polymérase III, appelés de type 1, 2 ou 3 (Schramm & Hernandez, 2002). Le seul promoteur de type 1 connu est celui du gène de l'ARNr 5S. Il s'agit d'un promoteur intragénique composé d'une boîte A, d'un élément intermédiaire (IE) et d'une boîte B qui constituent une région de contrôle interne (ICR). Les promoteurs de type 2, composés d'une boîte A et d'une boîte B très bien conservées, sont retrouvés chez tous les gènes d'ARN de transfert. Enfin, les promoteurs de type 3, les seuls à être extragéniques, sont caractéristiques du gène de l'ARN U6, mais peuvent également être retrouvés pour d'autres gènes chez les eucaryotes pluricellulaires. Chez *S. cerevisiae*, le gène *SNR6* présente une organisation hybride composée d'une boîte TATA en amont, d'une boîte B en aval et d'une boîte A interne (Figure 10).

## Introduction



**Figure 10. Les différents types de promoteurs de l'ARN polymérase III (d'après Schramm & Hernandez, 2002).**

Les promoteurs de type 1 sont constitués d'une région de contrôle interne (ICR) composée d'une boîte A, d'un élément intermédiaire (IE) et d'une boîte B. Les promoteurs de type 2 sont composés de boîtes A et B intragéniques. Les promoteurs de type 3 sont composés d'un élément de séquence proximal (PSE), d'un élément de séquence distal (DSE) et d'une boîte TATA. Chez *S. cerevisiae*, le promoteur du gène de l'ARN U6 est composé d'une boîte TATA, d'une boîte A et d'une boîte B.



**Figure 11. Recrutement de la machinerie de transcription par l'ARN polymérase III (d'après Schramm & Hernandez, 2002).**

Les flèches vertes représentent les interactions entre les protéines se liant à l'ADN et les séquences promotrices. Les flèches bleues représentent les interactions protéine-protéine entre les facteurs de transcription. Les flèches violettes représentent les interactions protéine-protéine entre les facteurs de transcription et l'ARN polymérase III.

**Tableau 5: Les facteurs généraux de transcription par l'ARN polymérase III chez *S. cerevisiae*.**

Facteur		Sous-unités	Gène
TFIIIA		TFIIIA	<i>TFC2</i>
TFIIB		TBP	<i>SPT15</i>
		Brf1	<i>BRF1</i>
		Bdp1 (B")	<i>TFC5</i>
TFIIC	$\tau$ A	$\tau$ 131	<i>TFC4</i>
		$\tau$ 95	<i>TFC1</i>
		$\tau$ 55	<i>TFC7</i>
	$\tau$ B	$\tau$ 138	<i>TFC3</i>
		$\tau$ 91	<i>TFC6</i>
		$\tau$ 60	<i>TFC8</i>

La machinerie de transcription par l'ARN polymérase III est composée de deux principaux facteurs, appelés TFIIB et TFIIC (Geiduschek & Kassavetis, 2001; Schramm & Hernandez, 2002) (Tableau 5). Chez *S. cerevisiae*, la purification du facteur TFIIB permet d'obtenir deux fractions d'activité différente, appelées B' et B". La fraction B' est constituée des protéines TBP et Brf1, alors que la fraction B" comporte entre autres la protéine Bdp1 (Kassavetis et al., 1995; R  th et al., 1996; Roberts et al., 1996). La formation du complexe de pr  initiation a lieu gr  ce    la reconnaissance des bo  tes A et B par le facteur TFIIC, permettant ainsi le recrutement du facteur TFIIB, puis de l'ARN polym  rase III. La transcription n'est pas initi  e de fa  on fid  le au niveau du site d'initiation correct en pr  sence de TFIIB seul. Celle-ci requiert l'implication d'autres prot  ines telles Nhp6a (Kassavetis & Steiner, 2006). Dans le cas de l'ARNr 5S, un troisi  me facteur appel   TFIIA est n  cessaire. Celui-ci est capable de reconnaître la r  gion ICR, facilitant ainsi le recrutement de TFIIC. Dans le cas des promoteurs de type 3 (absents chez *S. cerevisiae*), un quatri  me facteur appel   SNAPc est requis,    la place de TFIIC, pour la liaison    la s  quence PSE (Figure 11).

L'interaction entre TFIIB et l'ARN polym  rase III est rompue rapidement, apr  s l'incorporation des premiers ribonucl  otides (Kassavetis et al., 1992). De m  me que l'ARN polym  rase II, l'ARN polym  rase III peut subir plusieurs cycles d'initiation abortive avant d'entrer en   longation productive (Bhargava & Kassavetis, 1999). Bien que le facteur TFIIC (et TFIIA dans le cas de l'ARNr 5S) soit li      des promoteurs intrag  niques, celui-ci ne semble pas pr  senter un obstacle    l'  longation par l'ARN polym  rase III (Geiduschek & Kassavetis, 2001). L'  longation est un processus dynamique au cours duquel les ARN polym  rases peuvent rencontrer des sites de pause ou d'arr  t. *In vitro*, les sites de pause de

l'ARN polymérase III sont généralement localisés dans des régions riches en thymine. Le mécanisme de levée des blocages transcriptionnels est tout à fait similaire à celui de l'ARN polymérase II (voir aussi Chapitre II). L'activité endonucléase intrinsèque de l'ARN polymérase III est portée par les sous-unités Rpc160 et Rpc128 (Thuillier et al., 1996; Shaaban et al., 1996; Bobkova et al., 1999), et implique pour sa stimulation la sous-unité Rpc11, homologue à TFIIIS dans sa portion C-terminale (Chédin et al., 1998).

La terminaison de la transcription par l'ARN polymérase III a lieu grâce à la reconnaissance d'un signal de terminaison composé d'une succession de résidus thymine. Ce mécanisme nécessite la mise en jeu des sous-unités Rpc37 et Rpc53 de l'ARN polymérase III, qui en réduisant la vitesse d'élongation de l'ARN polymérase III, la rendrait plus apte à reconnaître les signaux de terminaison (Landrieux et al., 2006). Une importante caractéristique de l'ARN polymérase III est le couplage direct entre la terminaison et la réinitiation de la transcription. Le processus de réinitiation permet à l'ARN polymérase III de démarrer un nouveau cycle de transcription du même gène beaucoup plus rapidement, résultant ainsi en une meilleure efficacité de l'initiation (Dieci & Sentenac, 1996, 2003). La sous-unité Rpc11 de l'ARN polymérase III, en plus et de façon indépendante de son rôle dans l'activité de clivage, est indispensable à la réinitiation (Landrieux et al., 2006).

La régulation de la transcription par l'ARN polymérase III est encore relativement mal connue, bien qu'elle soit indispensable au bon fonctionnement de la cellule. Ainsi, chez les mammifères, la transcription par l'ARN polymérase III est réprimée par les suppresseurs de tumeurs Rb et p53 et au contraire stimulée par des oncogènes tels les protéines c-Myc ou ERK, suggérant un lien important entre la transcription par l'ARN polymérase III et la formation de cancers (White, 2004, 2008). Chez la levure *S. cerevisiae*, la protéine Maf1 a été identifiée comme un répresseur de la transcription par l'ARN polymérase III dans des conditions de stress ou de quiescence (Upadhyaya et al., 2002; Oficjalska-Pham et al., 2006; Cieśła & Boguta, 2008), dont la fonction est également conservée chez les eucaryotes pluricellulaires (Reina et al., 2006).

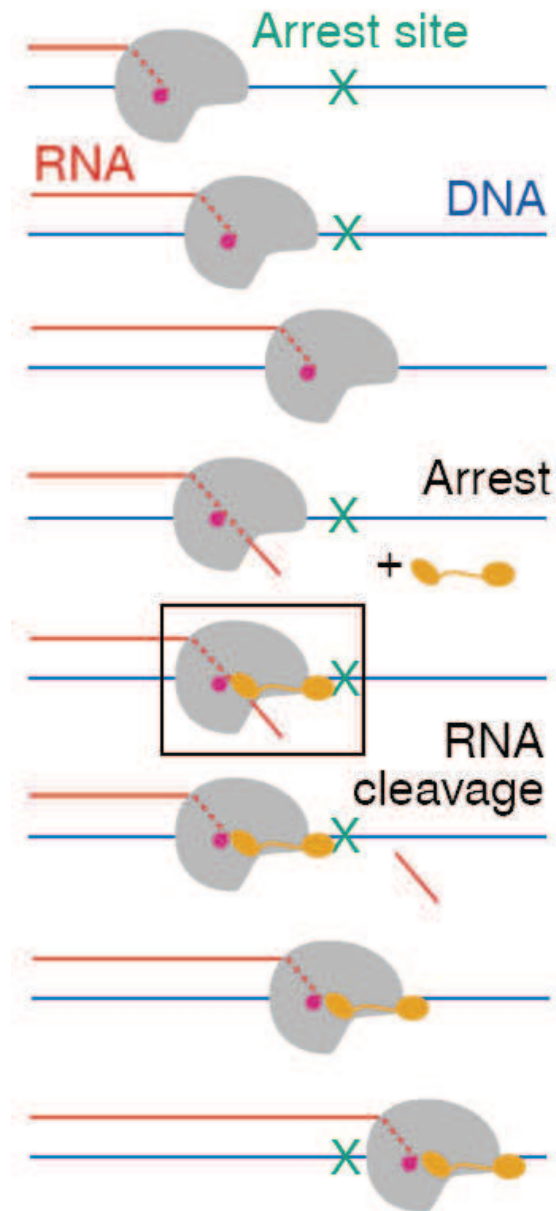
## CHAPITRE II

### *Le facteur de transcription TFIIS*

#### **a) Identification de TFIIS et de ses activités *in vitro***

TFIIS est un des premiers facteurs de transcription par l'ARN polymérase II à avoir été identifié. Il a été initialement décrit comme ayant une activité de stimulation non spécifique de l'ARN polymérase II purifiée à partir de cellules tumorales de souris (Natori et al., 1973; Sekimizu et al., 1976). Chez la levure *S. cerevisiae*, TFIIS a initialement été identifié comme un facteur, appelé P37 (37kDa), interagissant avec l'ARN polymérase II et stimulant son activité de transcription (Sawadogo et al., 1980). Il a été montré, chez différents organismes eucaryotes, que TFIIS agit après l'étape d'initiation, en facilitant le passage de l'ARN polymérase II à travers des sites de blocage, résultant ainsi en une augmentation de l'efficacité et de la processivité de la transcription par l'ARN polymérase II (Reinberg & Roeder, 1987; Rappaport et al., 1987; Reines et al., 1989; Sluder et al., 1989; Bengal et al., 1991). Certaines séquences d'ADN spécifiques qui bloquent l'ARN polymérase ont été identifiées (Bengal & Aloni, 1989; Wiest et al., 1992; Kash et al., 1993). Ces séquences sont en général riches en A-T ou caractérisées par une courbure de l'ADN (Kerppola & Kane, 1990; Fish & Kane, 2002). TFIIS stimule également l'élongation lorsque l'ARN polymérase rencontre des sites spécifiques de liaison de protéines à l'ADN ou des drogues liées à l'ADN (Reines et al., 1993; Reines & Mote, 1993; Mote et al., 1994).

TFIIS stimule donc le passage de l'ARN polymérase à travers des sites de blocage de l'élongation. L'ARN polymérase II ne progresse pas de façon uniforme le long de l'ADN, mais marque des pauses. Dans le cas d'une pause courte, l'ARN polymérase II reste compétente pour l'élongation, et l'ajout de nucléotides suffit à la reprise de l'élongation. Dans le cas d'une pause prolongée, l'ARN polymérase II devient bloquée en élongation, nécessitant la présence du facteur TFIIS pour stimuler la levée du blocage (Fish & Kane, 2002). Cette stimulation requiert le clivage de l'ARN néosynthétisé au sein du complexe ternaire d'élongation ARN polymérase II/ADN/ARN (Reines, 1992; Izban & Luse, 1992; Wang & Hawley, 1993). TFIIS réactive les complexes ternaires arrêtés selon un mécanisme qui implique le recul de l'ARN



**Figure 12. Modèle du passage par l'ARN polymérase II des sites d'arrêt grâce à l'activité de clivage de TFIIS (d'après Cramer, 2004).**

Lorsque l'ARN polymérase II est en cours d'élongation (gris), elle peut rencontrer un site d'arrêt (arrest site, vert) sur l'ADN (DNA, bleu). Si le blocage persiste, la pause est suivie d'un recul de l'ARN polymérase II qui désengage l'extrémité 3' de l'ARN (RNA, rouge) en cours de synthèse du site actif, provoquant l'arrêt de la transcription (arrest). TFIIS (orange) peut débloquer une ARN polymérase II arrêtée en stimulant le clivage du fragment d'ARN extrudé, permettant ainsi à l'ARN polymérase II de repartir en élongation.

polymérase II (Figure 12). Ce mouvement entraîne un décalage entre le site catalytique de l'ARN polymérase II et l'extrémité 3' de l'ARN en cours de synthèse, empêchant toute reprise ultérieure de l'élongation (Fish & Kane, 2002). TFIIS va alors stimuler une activité nucléase intrinsèque de l'ARN polymérase II, et permettre ainsi le clivage de l'ARN au niveau du site catalytique. La nouvelle extrémité 3' ainsi générée pourra alors servir de substrat pour la reprise de l'élongation (Gu et al., 1993; Rudd et al., 1994). La taille des produits de clivage peut varier d'un nucléotide à une dizaine de bases, les produits les plus courts étant généralement le résultat de pauses de courte durée dont le passage est indépendant de l'action de TFIIS (Izban & Luse, 1993; Gu & Reines, 1995). Il est possible que plusieurs cycles de blocage/clivage/reprise de la transcription soient nécessaires avant que l'ARN polymérase parvienne à traverser l'obstacle. L'ARN polymérase II pourra ainsi rencontrer de multiples fois un même site de blocage. L'efficacité de passage d'un site de blocage étant intrinsèque à chaque blocage, l'ARN polymérase II finira par le traverser si tant est que cela soit possible. En effet, TFIIS n'est pas capable de faciliter le passage à travers certains sites de blocage de l'ARN polymérase II tels les dimères de thymidine, pour lesquels la réparation du site est un préalable à toute reprise de l'élongation (Wind & Reines, 2000).

En plus des sites d'arrêt de la transcription préalablement répertoriés, TFIIS permettrait à l'ARN polymérase II de traverser les nucléosomes. En effet, les nucléosomes peuvent former une barrière particulièrement robuste, bloquant le passage de l'ARN polymérase II (Bondarenko et al., 2006). Des données obtenues *in vitro*, montrent que TFIIS et TFIIF permettent d'obtenir une vitesse d'élongation physiologique sur une matrice d'ADN nu. Cependant ces deux facteurs ne stimulent que légèrement l'efficacité de l'élongation par l'ARN polymérase II sur une matrice chromatinienne (Izban & Luse, 1992). Une étude plus récente, utilisant un complexe d'élongation de l'ARN polymérase II de *S. cerevisiae* artificiellement reconstitué sur une matrice d'ADN contenant un seul nucléosome, a montré que TFIIS accroît la transcription à travers la barrière nucléosomale, en induisant le clivage du transcrit dans les complexes arrêtés (Kireeva et al., 2005). La reconstitution d'un système de transcription par l'ARN polymérase II humaine hautement purifié sur une matrice contenant plusieurs nucléosomes, a confirmé que TFIIS, en synergie avec l'histone acétyltransférase p300, permet une élongation productive sur une matrice chromatinienne (Guermah et al., 2006, 2009). Par ailleurs, une étude récente de biophysique à molécule unique utilisant des complexes ternaires d'ARN polymérase II isolées *in vitro*, a montré que TFIIS augmente la limite de la force opérationnelle de l'ARN polymérase II, lui permettant ainsi de traverser des sites de blocage de la transcription (Galburt et al., 2007).



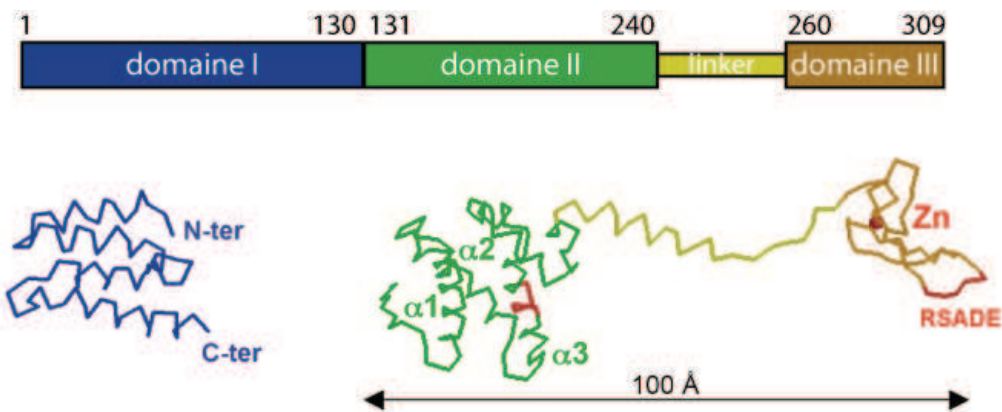
## **b) Structure et mode d'action de TFIIS *in vitro***

Le facteur TFIIS est composé de trois domaines structuraux distincts (Figure 13) (Morin et al., 1996): le domaine I composé des premiers 130 acides aminés de la protéine, le domaine II des acides aminés 131 à 240 et le domaine III des acides aminés 260 à 309. A ceux-ci s'ajoute un linker reliant les domaines II et III des acides aminés 241 à 259. La structure du TFIIS de *S. cerevisiae* a dans un premier temps été obtenue par RMN (Olmsted et al., 1998; Booth et al., 2000), puis par cristallographie à une résolution de 3,8 Å en complexe avec l'ARN polymérase II (Kettenberger et al., 2003) (Figure 13).

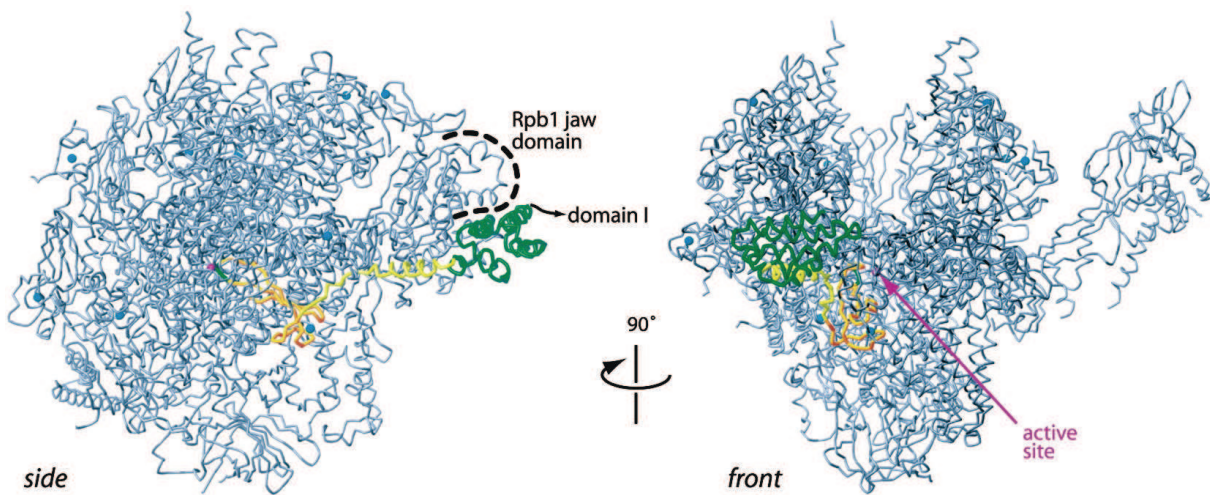
Le domaine I de TFIIS adopte une structure compacte de forme globulaire composée de quatre hélices  $\alpha$  structurellement indépendantes des domaines II et III (Booth et al., 2000). La fonction du domaine I reste encore inconnue, bien qu'il soit établi qu'il n'est pas essentiel à l'activité de clivage de TFIIS (Awrey et al., 1998). Le domaine I fusionné à la GST est suffisant pour purifier l'holoenzyme de l'ARN polymérase II humaine à partir de cellules HeLa (Pan et al., 1997). Cette holoenzyme contient, entre autres, l'ARN polymérase II ainsi que ses facteurs généraux de transcription TFIIB, une partie de TFIID, TFIIE, TFIIF, TFIIH, et la cycline Cdk8, une des sous-unités du Médiateur (homologue humain de Srb10 chez la levure). Elle permet l'initiation de la transcription *in vitro* en présence des activateurs spécifiques Gal4-VP16 et Gal4-Sp1. Chez la levure, le domaine I de TFIIS lie spécifiquement les sous-unités Med13 du Médiateur et Spt8 du coactivateur SAGA dans un test double hybride (Wery et al., 2004). Des expériences de co-immunoprécipitation ont confirmé que TFIIS peut se lier au module Cdk8 du Médiateur et à SAGA dans des extraits cellulaires. Ces données suggèrent un rôle de TFIIS dans l'initiation de la transcription, qui sera confirmé par la suite (voir ci-dessous).

Le domaine II de TFIIS est composé de trois hélices  $\alpha$  principales ( $\alpha$ 1-3) entourant un cœur hydrophobe (Morin et al., 1996; Olmsted et al., 1998) et de trois hélices  $\alpha$  plus petites ( $\alpha$ 4-6) qui ne se structurent qu'au contact de l'ARN polymérase II (Kettenberger et al., 2003). Les hélices  $\alpha$ 1 et  $\alpha$ 3 interagissent avec la sous-unité Rpb1 de la mâchoire de l'ARN polymérase II. L'hélice  $\alpha$ 3 du domaine II de TFIIS contient un motif basique dont certains résidus sont essentiels à l'interaction entre TFIIS et l'ARN polymérase II au niveau d'une boucle acide de la sous-unité Rpb1. De même, des mutations dans cette boucle acide réduisent drastiquement l'interaction de TFIIS avec l'ARN polymérase II (Wu et al., 1996). Le domaine

A



B



**Figure 13. Architecture de TFIIS (d'après Kettenberger et al., 2003).**

**A.** Structure et organisation en domaines de TFIIS. Le diagramme montre la répartition des trois domaines de TFIIS le long de la structure primaire de la séquence de la protéine. Les domaines I, II et III sont respectivement représentés en bleu, vert et marron. Le linker entre les domaines II et III est représenté en jaune. Les limites des domaines sont indiquées au dessus de la Figure en numérotant à partir du premier acide aminé N-terminal. **B.** Structure des résidus 148 à 309 de TFIIS liés à l'ARN polymérase II obtenue par diffraction aux rayons X vu de côté (side) ou de face (front). Le domaine II de TFIIS est lié à la mâchoire de Rpb1 (Rpb1 jaw domain). Le linker pénètre l'ARN polymérase et le domaine III est inséré directement au niveau du site actif (active site).

II de TFIIS interagit avec l'ARN polymérase II avec la même affinité que la protéine entière (Awrey et al., 1998), mais est incapable de stimuler l'activité de clivage.

Le domaine III de TFIIS est relié au domaine II par un linker non structuré lorsque la protéine est libre (Olmsted et al., 1998; Awrey et al., 1998), mais qui se structure en une grande hélice  $\alpha$  au contact de l'ARN polymérase II. Des mutations ponctuelles situées dans le linker abolissent l'activité de TFIIS, suggérant que cette région adopte une structure importante dans le complexe ternaire (Awrey et al., 1998).

Le domaine III de TFIIS adopte une structure de type ruban à zinc (zinc ribbon) constitué de trois feuillets  $\beta$  antiparallèles (Qian et al., 1993; Olmsted et al., 1998). Ces feuillets forment une boucle en épingle à cheveu acide qui atteint le site actif de l'ARN polymérase II (Kettenberger et al., 2003). Cette boucle contient un motif très conservé, le motif RSADE. Les deux résidus D290 et E291 sont essentiels pour l'activité de TFIIS, et des mutations de ces résidus rendent TFIIS totalement inactive sans pour autant altérer de façon importante la liaison à l'ARN polymérase II (Awrey et al., 1998). Le domaine III joue donc un rôle direct dans la catalyse du clivage du transcrit.

La structure cristalline de l'ARN polymérase II interagissant avec TFIIS a permis de mieux comprendre le mode d'action de TFIIS en tant que facteur de clivage (Kettenberger et al., 2003). TFIIS ouvre une crevasse dans l'ARN polymérase II via le linker, permettant ainsi au domaine III de s'insérer profondément dans le pore, à proximité immédiate du site actif. Le motif RSADE, et en particulier les résidus D et E vont alors pouvoir compléter le site catalytique de l'ARN polymérase II et positionner un ion métallique de magnésium et une molécule d'eau pour permettre le clivage. L'interaction avec TFIIS induit de profonds changements structuraux de l'ARN polymérase II, et en particulier de son site actif, la rendant ainsi compétente pour le clivage. Par ailleurs, la liaison de TFIIS réaligne l'ARN néosynthétisé au niveau de site actif, rendant ainsi probablement le complexe d'élongation moins sensible aux sites de pause (Kettenberger et al., 2004). Une structure de l'ARN polymérase II en situation de recul à une résolution de 3,4 Å a été obtenue récemment (Wang et al., 2009). Cette structure révèle que le recul de l'ARN polymérase II d'un seul nucléotide constitue un état stable, favorisant le clivage et le relarguage d'un dinucléotide en présence de TFIIS. L'ARN polymérase II en situation de recul présente un site caractéristique, appelé site "P" capable de se lier à ce nucléotide en cas de mauvaise incorporation d'un nucléotide ou d'un dommage à l'ADN. L'activité de clivage de l'ARN polymérase II serait ainsi partie intégrante de sa capacité de fidélité.

### c) Fonction de TFIIS

Bien que le rôle de TFIIS *in vitro* ait été abondamment étudié depuis plus de trente ans, sa fonction *in vivo* reste encore mal comprise. Le facteur TFIIS de *S. cerevisiae* a été purifié (Nakanishi et al., 1992), et il a été montré que cette protéine était issue de la transcription du gène *PPR2* (pyrimidine pathway regulator). *PPR2* a été identifié dans un criblage génétique recherchant des mutants de levure sensibles à un analogue de nucléotide, le 6-azauracil (6AU) (Hubert et al., 1983). Le nom usuel actuel du gène codant TFIIS est *DSTI* (DNA strand transferase), car TFIIS a également été identifié comme participant au transfert de brin lors de la recombinaison homologue méiotique (Clark et al., 1991), bien que depuis rien n'ait confirmé cette hypothèse. L'inactivation de TFIIS chez la levure n'a que très peu d'effet sur la croissance cellulaire. Cependant elle provoque une sensibilité à certaines drogues telles le 6AU, un inhibiteur de l'orotidylate décarboxylase (Ura2) et de l'IMP déhydrogénase (Imd2), ou le mycophénolate (MPA), un inhibiteur de l'IMP déhydrogénase. Ces deux drogues entraînent une diminution de la concentration cellulaire en GTP et/ou UTP (Exinger & Lacroute, 1992; Nakanishi et al., 1992), suggérant qu'une quantité limitante de nucléotides dans la cellule freine l'ARN polymérase II en cours d'élongation, rendant ainsi TFIIS essentiel. Cependant, des expériences d'immunoprécipitation de la chromatine ont montré qu'en présence de 6AU, la délétion de *DSTI* affecte la processivité de l'ARN polymérase II, mais pas sa vitesse d'élongation (Mason & Struhl, 2005). Enfin, TFIIS permet d'augmenter la fidélité de la transcription par l'ARN polymérase II, bien que ce rôle de TFIIS n'ait pas clairement été mis en évidence *in vivo* (Jeon & Agarwal, 1996; Thomas et al., 1998; Koyama et al., 2007). L'inactivation de *DSTI* peut être complétée par les domaines II et III de TFIIS, mais pas par des mutants du motif RSADE (Ubukata et al., 2003), confirmant ainsi l'importance de ce motif pour la fonction de TFIIS *in vivo*.

La construction d'un site d'arrêt artificiel (ARTAR) de la transcription a confirmé la capacité de TFIIS à stimuler l'élongation de la transcription à travers un site d'arrêt *in vivo* (Kulish & Struhl, 2001). Cette stimulation semble être principalement due à l'activité de clivage de TFIIS plutôt qu'à son activité de stimulation du passage des sites d'arrêt. En effet, des mutants actifs pour le clivage mais pas pour le passage des sites d'arrêt présentent le même phénotype qu'une souche sauvage (Ubukata et al., 2003). Cette observation suggère, par ailleurs, que d'autres facteurs permettraient spécifiquement à l'ARN polymérase II de passer à travers des sites d'arrêt.

La fonction de TFIIIS *in vivo* a principalement été étudiée chez la levure *S. cerevisiae*. Cependant, une étude récente a mis en évidence que la diminution de l'expression du gène codant TFIIIS réduit significativement la prolifération des cellules cancéreuses dans des lignées cellulaires de cancer du sein, des poumons et du pancréas (Hubbard et al., 2008). La diminution de l'expression du gène codant TFIIIS dans des lignées cellulaires de cancer du sein (MCF7) induit la mort des cellules cancéreuses et augmente l'expression de c-myc et p53, alors que des cellules non-cancéreuses (MCF10A) sont moins affectées. De plus, il a été montré que, *in vivo*, une diminution de la transcription de TFIIIS peut être tolérée au niveau cellulaire et physiologique. Chez la souris, TFIIIS a un rôle critique dans l'hématopoïèse. En effet, chez des souris "knock-out" les embryons ne peuvent atteindre leur terme pour cause d'anémie, bien que la croissance cellulaire, la différenciation ni le développement des embryons ne soient altérés jusqu'au stade E13.5 (Ito et al., 2006).

Chez la drosophile, les gènes de choc thermique ne sont que très faiblement transcrits dans les conditions normales, mais l'ARN polymérase II est séquestrée au promoteur, en état de pause, pour permettre une réponse rapide dès l'induction de la transcription. La présence de TFIIIS est essentielle à l'induction rapide de la transcription du gène *Hsp70*, car il permet à la polymérase de sortir de son état de pause au promoteur (Adelman et al., 2005).

#### **d) Le clivage du transcrit : une fonction conservée**

L'activité de clivage du transcrit est une fonction extrêmement bien conservée au cours de l'évolution. Elle est générale à toutes les ARN polymérases ADN-dépendantes. Certaines ARN polymérases sont capables de cliver le transcrit sans l'aide d'une protéine extérieure, alors que d'autres nécessitent la mise en jeu d'un facteur capable de réactiver la molécule d'ARN polymérase arrêtée. Chez les procaryotes, les protéines GreA et GreB stimulent l'activité de clivage de l'ARN polymérase bactérienne. Les protéines GreA et GreB d'*Escherichia coli* sont homologues entre elles (Borukhov et al., 1993), et possèdent des homologues dans tous les génomes bactériens. La résolution de la structure cristalline de GreA révèle une extrémité N-terminale super-enroulée (coiled coil) interagissant avec l'extrémité N-terminale du transcrit et stimulant la réaction de clivage (Stebbins et al., 1995; Koulich et al., 1997). La structure de GreB interagissant avec l'ARN polymérase bactérienne montre que l'extrémité N-terminale de cette protéine pénètre profondément dans l'ARN

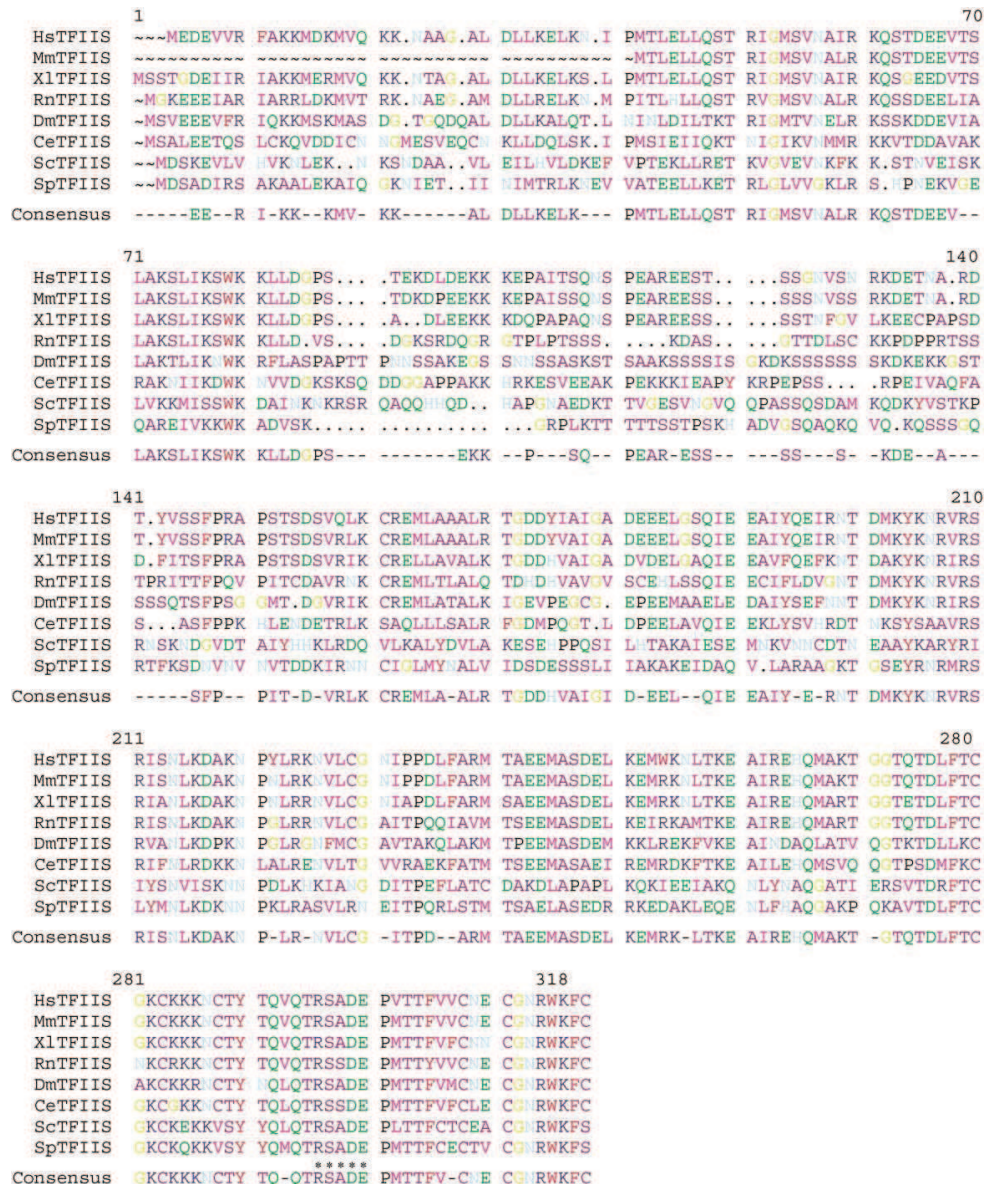
polymérase jusqu'à son site actif (Opalka et al., 2003). GreA et GreB stimulent le clivage du transcrit de deux façons différentes. GreA induit le clivage de fragments d'ARN de 2-3 nucléotides, et peut uniquement prévenir la formation de complexes d'élongation arrêtés, alors que GreB induit le clivage de fragments pouvant aller jusqu'à 18 nucléotides et peut débloquent des complexes d'élongation précédemment arrêtés. (Borukhov et al., 1993). En dépit de leur fonction et de leur mécanisme moléculaire très similaire, les facteurs GreA/B procaryotiques et le facteur TFIIS eucaryotique n'ont aucune similarité de séquence ou de structure et ont probablement évolué indépendamment. Le clivage du transcrit semble donc être une activité particulièrement importante. Une activité similaire a été rapportée pour l'ARN polymérase du virus de la vaccine (Hagler & Shuman, 1993), l'ARN polymérase monomérique du bactériophage T7 (Sastry & Ross, 1997), ainsi que l'ARN polymérase des Archaea pour laquelle un facteur homologue à TFIIS a été identifié (Hausner et al., 2000).

Chez les eucaryotes, la protéine TFIIS peut être présente en une ou plusieurs isoformes. Chez *S. cerevisiae*, *Schizosaccharomyces pombe*, *Drosophila melanogaster* et *Caenorhabditis elegans* un seul gène a été identifié. TFIIS n'a été que très peu étudié chez les plantes, et son rôle dans la stimulation du clivage du transcrit n'a jamais été caractérisé. Cependant, un homologue de TFIIS, complétant partiellement la sensibilité au 6AU de la délétion de *DST1* chez *S. cerevisiae*, a été identifié récemment dans le génome d'*Arabidopsis thaliana*. Les auteurs ont montré que TFIIS jouerait un rôle critique dans le développement de la graine (Grasser et al., 2009). Chez le xénope, le rat, la souris et l'homme, les gènes codant TFIIS forment une famille dont les membres présentent une expression différentielle selon les tissus (Labhart & Morgan, 1998). L'une des isoformes, généralement exprimée dans tous les tissus, est appelée la forme générale de TFIIS (Ito et al., 2000). La seconde isoforme est principalement retrouvée dans les testicules et les ovaires (Xu et al., 1994; Umehara et al., 1995; Ito et al., 1996; Weaver & Kane, 1997), alors que la troisième a été retrouvée principalement exprimée dans le cœur, le foie, le muscle squelettique et le rein (Taira et al., 1998). Ces trois gènes ont été nommés *Tcea1*, *Tcea2* et *Tcea3*. Dans la plupart des cas, les régions N- et C-terminales de TFIIS sont très fortement conservées, alors que la région reliant ces deux domaines varie considérablement du point de vue de sa séquence et de sa longueur (Figure 14) (Labhart & Morgan, 1998).

Malgré son excellente conservation au cours de l'évolution, aussi bien en terme de séquence que de fonction, la délétion de TFIIS n'est pas létale chez *S. cerevisiae*, ou *S. pombe* en conditions standard de croissance, alors même que ce gène n'existe que sous une seule copie chez ces organismes. Il devient cependant essentiel dans des conditions particulières



## Introduction



**Figure 14. Alignement des séquences des protéines TFIIS de différents organismes (d'après Fish & Kane, 2002).**

Les formes générales des protéines TFIIS de différents organismes sont alignées en utilisant une matrice Blosom62 et le programme GCG. Une séquence consensus a été déterminée si un acide aminé est conservé dans au moins quatre organismes. Les organismes représentés sont: *Homo sapiens* (Hs), *Mus musculus* (Mm), *Xenopus laevis* (Xl), *Rattus norvegicus* (Rn), *Drosophila melanogaster* (Dm), *Caenorhabditis elegans* (Ce), *Saccharomyces cerevisiae* (Sc), et *Schizosaccharomyces pombe* (Sp). Les acides aminés, représentés par leur code en une lettre sont colorés de la façon suivante: I, V, L, M = magenta; G = orange; P = noir; Q = vert; F, Y, W = rouge; H, N = bleu clair; D, E = vert foncé; K, R = bleu; A, C, S, T = violet. Le motif RSADE est surmonté par des astérisques.

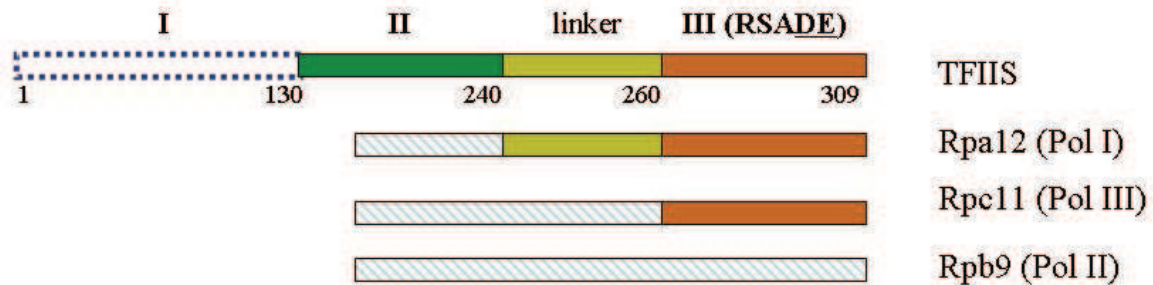
telles que la carence en nucléotides. Ce paradoxe peut être expliqué par la différence de taille des gènes entre les espèces. En effet, les génomes de levures étant plus petits que ceux des eucaryotes pluricellulaires, notamment du fait de leurs gènes plus courts et généralement dépourvus d'introns, il est possible que la transcription des gènes soit moins affectée par l'absence de TFIIS, chaque gène ayant un plus faible nombre de sites d'arrêt potentiels. Par ailleurs, il est possible que chez la levure, d'autres facteurs puissent partiellement corriger la délétion de TFIIS. Il a ainsi été montré que la délétion de *dst1* est colétale avec celle de facteurs de remodelage de la chromatine ou de facteurs de transcription (Davie & Kane, 2000; Fish et al., 2006). Ceci suggère qu'il pourrait exister une certaine redondance fonctionnelle entre différents acteurs de la transcription.

Chez la levure *S. cerevisiae*, l'activité de clivage du transcrit a d'abord été découverte pour l'ARN polymérase II (Reines, 1992). Une activité similaire a également été reportée pour l'ARN polymérase III (Whitehall et al., 1994) et I (Tschochner, 1996). Celle-ci dépend des sous-unités Rpc11 (Chédin et al., 1998) et Rpa12 (Kuhn et al., 2007) de l'ARN polymérase III et I, respectivement. Il est à noter que des travaux antérieurs, qui n'ont pas été poursuivis, avaient suggéré que TFIIS pourrait être impliqué dans l'activité de clivage de l'ARN polymérase I (Sawadogo et al., 1980, 1981; Schnapp et al., 1996). Les domaines C-terminaux de Rpa12 et Rpc11 possèdent, comme TFIIS, le motif RSADE indispensable à l'activité de clivage ainsi que deux domaines de liaison au zinc. Rpb9, paralogue de Rpa12 et Rpc11 dans l'ARN polymérase II n'a cependant pas de motif RSADE, mais coopère avec TFIIS pour assurer son activité de clivage (Awrey et al., 1997) (Figure 15). En effet, une ARN polymérase II dépourvue de la sous-unité Rpb9 est moins sujette aux pauses que l'ARN polymérase sauvage. Par ailleurs, les complexes d'ARN polymérase II Rpb9 $\Delta$  arrêtés sont moins sensibles à l'action de TFIIS que le sauvage. Ces données suggèrent que la sous-unité Rpb9 transmet un signal à TFIIS, induisant ainsi un changement de conformation de l'ARN polymérase II.

Par ailleurs, les domaines I et II de TFIIS, bien qu'ils ne soient pas directement impliqués dans sa fonction de clivage du transcrit, sont également bien conservés. Le domaine I de TFIIS est conservé de la levure à l'homme, et est homologue à une région de l'élongine A, un autre facteur d'élongation de transcription (Shilatifard et al., 1996), et de la sous-unité Med26 du Médiateur (CRSP70) (Ryu et al., 1999). La résolution de la structure du domaine I de TFIIS par RMN a permis de modéliser celle des domaines N-terminaux de ces deux protéines (Booth et al., 2000). Le domaine II de TFIIS présente une homologie avec le domaine central de la protéine Bye1 chez *S. cerevisiae*. Bye1 est l'homologue de PHF3 chez



l'homme, une protéine qui serait impliquée dans la formation des glioblastomes (Fischer et al., 2001; Struss et al., 2001), et pourrait être un compétiteur spécifique de TFIIS pour sa liaison à l'ARN polymérase II.



**Figure 15: Conservation des domaines de TFIIS chez les protéines Rpa12 (ARN polymérase I), Rpc11 (ARN polymérase III) et Rpb9 (ARN polymérase II).**

Les domaines I est représenté en pointillés bleus. Le domaine II, le linker et le domaine III sont respectivement représentés vert, jaune et orange.

### e) Rôle de TFIIS dans l'initiation de la transcription

Le caractère non-essentiel de la délétion de *DST1* chez la levure *S. cerevisiae* a encouragé de nombreux auteurs à utiliser des cribles de létalité synthétique afin de mieux caractériser le rôle de TFIIS *in vivo*, mais également d'identifier de nouveaux facteurs impliqués dans la régulation de l'élongation de la transcription par l'ARN polymérase II (Davie & Kane, 2000). Cependant, et de façon inattendue, ces cribles ont également permis de suggérer un nouveau rôle de TFIIS dans l'initiation de la transcription par l'ARN polymérase II. En effet, l'inactivation de *DST1* est colétale avec celle de *MED31*, une des sous-unités du Médiateur de la transcription. Cette colétalité peut être supprimée par la surexpression de gènes impliqués dans l'initiation de la transcription tels que *MED15*, une autre sous-unité du Médiateur, *SUA7*, le gène codant le facteur général de transcription TFIIB, ou *TAF14*, codant un facteur faisant partie de différents complexes dont TFIID et SWI/SNF (Malagon et al., 2004). D'autres données sont en accord avec un rôle de TFIIS dans l'initiation de la transcription. Tout d'abord, le domaine I de TFIIS interagit physiquement avec des sous-unités des coactivateurs SAGA et Médiateur (Wery et al., 2004). *In vitro*, TFIIS peut partiellement compenser la perte de la sous-unité Med31 et de la partie N-terminale de la sous-unité Med7 du Médiateur (Koschubs et al., 2009). De plus, des expériences

d'immunoprécipitation de la chromatine ont révélé la présence de TFIIS au niveau de l'UAS du gène *GALI* et ont montré que le recrutement de TFIIS est dépendant de la présence de SAGA et du Médiateur, mais indépendant du recrutement de l'ARN polymérase II (Prather et al., 2005). Le rôle de TFIIS dans l'initiation de la transcription requiert, *in vivo*, son domaine de liaison à l'ARN polymérase II, mais est indépendant de son activité de clivage (Guglielmi et al., 2007). Enfin, il a été montré par spectrométrie de masse que TFIIS fait partie du complexe de préinitiation de l'ARN polymérase II. Les domaines I et II de TFIIS sont essentiels à la formation du complexe de préinitiation *in vitro* (Kim et al., 2007). Toutes ces données suggèrent un rôle important de TFIIS dans l'initiation de la transcription. Le modèle actuel propose que TFIIS soit recruté par des activateurs ou des coactivateurs de façon indépendante de l'ARN polymérase II au niveau des promoteurs de certains gènes. TFIIS pourrait alors, en conjonction avec des coactivateurs tels que le Médiateur, stimuler le recrutement ou la stabilisation de l'ARN polymérase II sur les promoteurs en interagissant directement avec elle au niveau de son domaine II (Guglielmi et al., 2007).

## CHAPITRE III

### *Exploration fonctionnelle des génomes: de la levure à l'homme*

#### **a) Le séquençage des génomes eucaryotes: utilité de *S. cerevisiae* comme modèle**

*S. cerevisiae* est un organisme modèle idéal pour les analyses faisant appel à la génétique, la biologie moléculaire ou la biochimie. Comparé aux autres eucaryotes, *S. cerevisiae* a un génome particulièrement compact. Environ 70% de l'ADN code des protéines (déduction faite de l'ADNr) dont la très grande majorité est dépourvue d'introns. *S. cerevisiae* peut être cultivée facilement et rapidement aussi bien sous forme haploïde que diploïde, rendant l'analyse génétique particulièrement aisée. En effet, des mutations récessives peuvent être facilement obtenues en travaillant avec des cellules haploïdes et la complémentation génétique ne nécessite que l'association de deux formes haploïdes pour obtenir un organisme diploïde. De plus, *S. cerevisiae* peut être facilement génétiquement modifiée par transformation d'ADN et recombinaison homologue. Ces techniques ont été rendues possibles grâce à l'utilisation de marqueurs génétiques tels que les marqueurs d'auxotrophie qui permettent une sélection efficace, et de vecteurs navettes qui l'on peut amplifier et manipuler dans la bactérie puis transformer dans la levure.

La plupart des processus cellulaires présents chez la levure *S. cerevisiae* sont conservés chez les eucaryotes pluricellulaires, ce qui en fait un bon modèle pour l'étude des maladies héréditaires, comme par exemple celles associées à des mutations de l'ADN mitochondrial (Barrientos, 2003), mais aussi pour la recherche en oncologie (Kurtz et al., 2004) ou l'étude des prions (Wickner et al., 2007). Pour toutes ces raisons, *S. cerevisiae* est, depuis plus d'un demi siècle, utilisée comme un organisme modèle de choix pour l'étude des eucaryotes avec, depuis le séquençage de son génome en 1996, de nombreuses études pionnières de génomique fonctionnelle.

La levure *S. cerevisiae* a été le premier eucaryote dont le génome a été entièrement séquencé (Goffeau et al., 1996). Son génome est relativement petit, comportant un total d'environ 12 mégabases réparties sur 16 chromosomes auxquels il faut ajouter 1,2 mégabases correspondant à l'ADNr, fait de quelques 200 répétitions en tandem localisées sur le chromosome XII. Les bases de données de référence (*Saccharomyces* Genome Database

(SGD, <http://www.yeastgenome.org/>), YPD (<https://portal.biobase-international.com/>)) recensent aujourd'hui environ 5800 ORFs confirmés dont environ 80% sont caractérisées. L'excellente annotation du génome de la levure tient, entre autres, à l'effort constant de séquençage entrepris par la communauté levuriste, notamment à travers le projet Génolevure (<http://www.genolevures.org/>), un projet de génomique comparée chez les levures hémiascomycètes. L'une des grandes surprises du séquençage de *S. cerevisiae* a été la découverte de très nombreuses duplications segmentales (Wolfe & Shields, 1997; Kellis et al., 2004). Celles-ci sont dues à une duplication ancestrale du génome après la divergence entre le genre *Saccharomyces* et le genre *Kluyveromyces*, suivie d'une perte progressive d'une partie de l'équipement génétique suite à des délétions.

L'intérêt porté à l'étude fonctionnelle chez la levure a permis l'émergence de nombreuses approches d'analyses globales ayant amélioré l'annotation fonctionnelle du génome (Kumar & Snyder, 2001; Grünenfelder & Winzeler, 2002). L'analyse du transcriptome en est l'une des principales, et sera développé dans les paragraphes suivants. L'une des autres stratégies pour mieux comprendre la fonction d'un gène est d'étudier les conséquences fonctionnelles de sa délétion. Plusieurs méthodes complémentaires ont été décrites: production et caractérisation phénotypique de collections de mutants issus de mutagenèse aléatoire par étiquetage de transposons (Ross-Macdonald et al., 1999), ou de délétion systématique de chaque ORF du génome (Winzeler et al., 1999; Giaever et al., 2002). Ces études ont révélé que seul 18% des gènes de *S. cerevisiae* sont essentiels à sa croissance sur milieu riche (glucose). Le nombre inattendu de gènes dont la délétion n'a aucun effet apparent suggère la présence de mécanismes évolutifs compensant les effets délétères d'une perte de gène. L'une des hypothèses avancées repose sur la duplication ancestrale du génome de *S. cerevisiae*, qui aurait créé des paires de gènes dupliqués dont la fonction a légèrement divergé, mais qui restent capables de corriger les défauts engendrés par la délétion de l'un des partenaires (Stein & Aloy, 2008). Une extension de ce type d'analyse consiste en la délétion de deux ou plusieurs gènes en même temps afin de rechercher une interaction génétique entre ces gènes. La recherche de mutants synthétiques létaux est l'une des stratégies les plus utilisées, et a permis d'identifier des complexes fonctionnels multi-protéiques (Collins et al., 2007).

La compréhension des liens fonctionnels entre les protéines peut aussi être explorée par l'identification d'interactions physiques entre les protéines, par exemple grâce à l'utilisation du "double-hybride" à grande échelle (Uetz et al., 2000; Walhout et al., 2000; Ito et al., 2001) ou de la spectrométrie de masse (Gavin et al., 2002; Ho et al., 2002). Le rôle d'un

grand nombre de protéines nucléaires peut également être étudié par l'identification de leur localisation globale sur la chromatine grâce à la technique de "ChIP-on-chip" (Ren et al., 2000; Iyer et al., 2001). Enfin, d'autres travaux ont porté sur l'étude de l'expression et de la localisation des protéines grâce à leur fusion à une étiquette de type TAP ou GFP (Ghaemmaghami et al., 2003; Huh et al., 2003). Au cours de ma thèse, je me suis plus particulièrement intéressée à la technologie des puces à ADN, et à son utilisation pour l'étude globale de la transcription.

Après une présentation générale de la technique des puces à ADN, je m'attacherai à décrire différents travaux fondateurs pour l'étude du transcriptome chez la levure. Je présenterais ensuite la technique d'immunoprécipitation de la chromatine pour l'étude de la distribution des protéines sur le génome et différents domaines d'applications courantes. Pour finir, je décrirai les avancées du séquençage à haut débit et les apports de cette technologie pour l'étude de la transcription.

## **b) Les puces à ADN: présentation générale de la technique**

Aujourd'hui les puces à ADN sont utilisées dans des domaines aussi variés que l'analyse du transcriptome, l'étude de la localisation des protéines sur l'ADN (ChIP-on-chip), la variation du nombre de copies d'ADN (CGH, "Comparative Genomic Hybridization") ou l'étude du polymorphisme (SNP, "Single Nucleotide Polymorphism"). La technologie des puces à ADN a profondément transformé la biologie moléculaire au cours de la dernière décennie. Elle n'a cessé d'évoluer et a participé à l'essor d'une nouvelle discipline: la génomique fonctionnelle. La génomique fonctionnelle, l'étape suivante au séquençage d'un génome, repose sur la caractérisation fonctionnelle systématique des gènes. Elle permet la compréhension du fonctionnement des gènes et des autres composantes du génome, leur expression et leur régulation, et est indispensable à la prédiction du fonctionnement d'un organisme.

La technique des puces à ADN repose sur la propriété de deux brins d'ADN à pouvoir être séparés et réassociés grâce à des variations de température (Marmur & Doty, 1961). Cette propriété a été utilisée de façon ingénieuse par E. Southern, l'inventeur du "Southern blot", une technique de base de la biologie moléculaire qui permet de détecter et quantifier la présence d'une séquence nucléique spécifique au sein d'un échantillon biologique complexe, par hybridation d'une sonde de séquence complémentaire portant un marquage radioactif (Southern, 1975). Lors d'un colloque en 1989, E. Southern a présenté une méthode de

fabrication de réseaux d'oligonucléotides fixés sur lame de verre qui, en principe, permettrait de déduire les séquences de brins d'ADN inconnues d'après leur hybridation avec des fragments d'ADN sonde connus. A la même période, S. Fodor parvient à développer une technique permettant de synthétiser *in situ* des oligonucléotides sur lame de verre par photolithographie (Fodor et al., 1991). Quelques années plus tard, S. Fodor créera la société Affymetrix.

A la fin des années 1980, les premiers "filtres à haute densité" (aussi appelés "macroarray" par opposition aux "microarray"), des membranes en nylon portant plusieurs milliers de clones d'ADN, font leur apparition (Craig et al., 1990). Elles sont progressivement supplantées par les puces sur lame de verre (Schena et al., 1995) (ou "microarray") à partir du milieu des années 1990. Très rapidement, une première puce contenant toutes les ORFs de la levure *S. cerevisiae* a été développée (Lashkari et al., 1997). Il faudra attendre la fin des années 1990 pour voir apparaître les premiers logiciels pour l'analyse statistique des données de puces à ADN (Eisen et al., 1998). Les détails de l'approche utilisée pour l'analyse statistique des données de ChIP-on-chip, sont présentés en Annexe.

Une puce est généralement définie comme une collection de quelques centaines à plusieurs dizaines de milliers de sondes ou "spots" (de l'anglais spot = tache), chacune étant constituée d'un dépôt de fragments d'ADN ou d'oligonucléotides, liés de façon ordonnée à un support solide (généralement une lame de verre). L'hybridation de la puce avec un échantillon biologique, marqué par une molécule fluorescente, permet de détecter et de quantifier l'ensemble des cibles qu'il contient en une seule expérience. Cette approche a profondément modifié notre façon de concevoir la recherche en biologie, et a permis le passage d'études ciblées sur un seul ou quelques gènes à des études portant sur l'ensemble du génome.

Les premières puces à ADN, appelées puces "spottées", ont été développées par P. Brown et coll (Schena et al., 1995). Des solutions d'ADN sont préparées soit par amplification PCR à partir du génome ou de banques d'ADN complémentaires, soit par synthèse d'oligonucléotides. Des micro-gouttelettes de ces solutions sont ensuite déposées par un robot, selon une matrice d'emplacements définis, sur une lame de verre traitée par un revêtement chimique qui permet de fixer l'ADN. Au cours de la dernière décennie, les progrès de la robotique ont conduit à la miniaturisation des puces à ADN, et au développement de différentes approches de fabrication. La densité des spots, la méthode de dépôt sur la lame, la longueur des sondes et le format de la puce sont les principaux facteurs qui différencient les plateformes existantes à l'heure actuelle. Les caractéristiques des puces les plus courantes sont résumées ci-dessous:

- Les puces Agilent

Méthode de dépôt: Les puces Agilent sont synthétisées par la méthode "jet d'encre", qui est adaptée de la technologie utilisée par les imprimantes jet d'encre. Le système Rosetta utilise directement des cartouches légèrement modifiées fournies par les imprimantes Epson. Grâce à cette méthodologie, des oligonucléotides sont synthétisés "goutte à goutte" directement sur la lame de façon extrêmement précise en terme de quantité déposée et de régularité des dépôts (Figure 16).

Longueur des sondes: 60-mer.

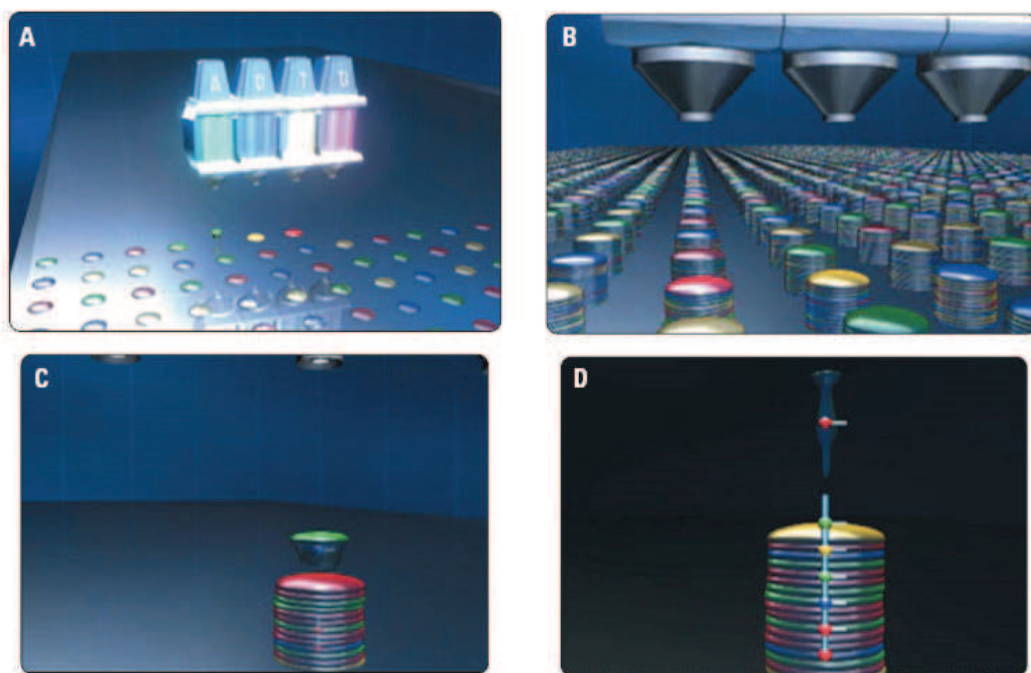
Format: 244000 spots par puce.

- Les puces Nimblegen

Méthode de dépôt: Les puces Nimblegen sont synthétisées grâce à un procédé appelé "Maskless Array Synthesizer" (MAS). Au cœur du système se trouve un "Digital Micromirror Device" (DMD) composé de milliers de micro-miroirs indépendants les uns des autres (Figure 17). Le DMD crée des "masques virtuels" qui remplacent les masques physiques utilisés pour les puces traditionnelles (voir puces Affymetrix). Ces "masques virtuels" reflètent la lumière UV selon un schéma établi grâce aux miroirs contrôlés par un ordinateur. La lumière UV clive de façon spécifique un groupement protecteur photolabile à la position exacte où le prochain nucléotide sera inséré (Figure 18).

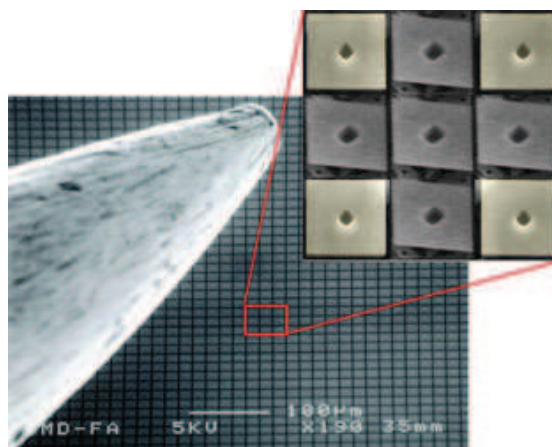
Longueur des sondes: 50 à 85-mer.

Format: 2,1 million de spots par puce.



**Figure 16.** Schéma de la synthèse des oligonucléotides chez Agilent (d'après <http://www.chem.agilent.com/en-US/Products/Instruments/dnamicroarrays/Pages/gp557.aspx>).

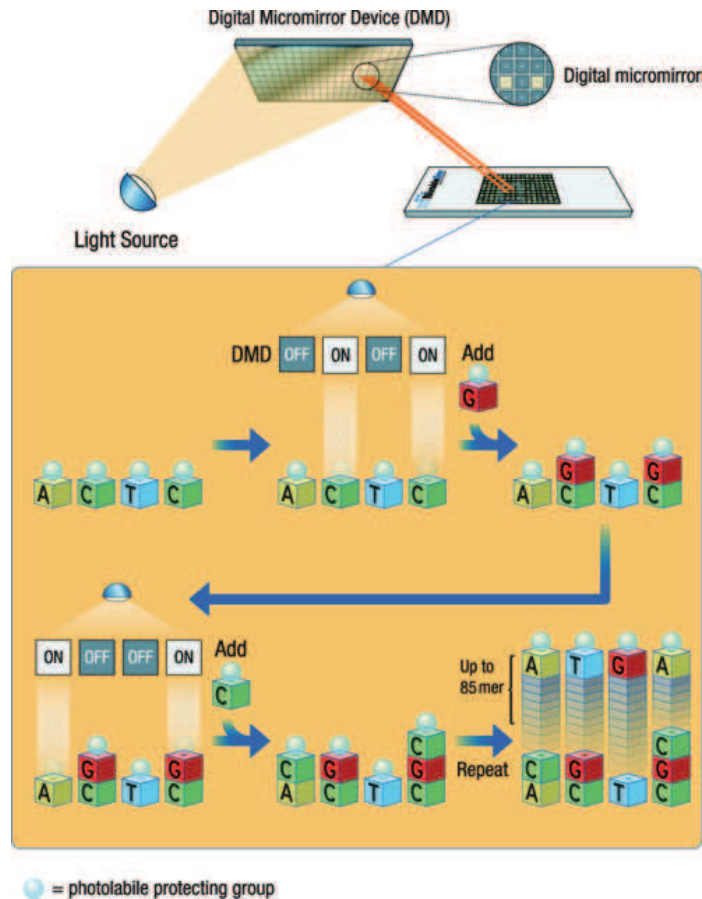
Ces quatre images illustrent le mécanisme général de la synthèse des oligonucléotides par impression jet d'encre. **A.** la première couche de nucléotides est déposée sur la surface de la puce à ADN. **B.** La croissance des oligonucléotides est représentée après que plusieurs couches de nucléotides aient été imprimées précisément. **C et D.** Vue détaillée d'un oligonucléotide lorsqu'une nouvelle base est ajoutée à la chaîne.



**Figure 17.** Les micro-miroirs utilisées par Nimblegen (d'après <http://www.nimblegen.com/technology/manufacture.html>).

Les miroirs du "Digital Micromirror Device" (DMD) sont représentés en comparaison avec la pointe d'une épingle.





**Figure 18.** Schéma de synthèse *in situ* des oligonucléotides chez Nimblegen. (d'après <http://www.nimblegen.com/technology/manufacture.html>).

Les micro-miroirs réfléchissent la lumière UV, déprotégeant ainsi l'oligonucléotide et permettant l'ajout de la base suivante.

#### - Les puces Affymetrix

Méthode de dépôt: Le principe de fabrication des puces à ADN Affymetrix (appelées GeneChip) est la synthèse *in situ* de centaines de milliers de séquences distinctes d'oligonucléotides sur une matrice de verre, fixées par photolithographie et chimie combinatoire (Figure 19). La photolithographie consiste en l'utilisation de masque photosensible. Ainsi au départ, tous les sites de fixation des sondes sont protégés par un groupement photolabile. L'éclairage des régions spécifiques (définies par le masque) du support en verre permet d'enlever le groupement photolabile et ainsi de fixer un nucléotide possédant lui aussi un groupement photolabile. Chaque oligonucléotide de 25 bases est reproduit à des millions d'exemplaires dans une région circonscrite. Le principe des puces Affymetrix repose sur l'existence de séquences dites consensus servant de références

spécifiques à un gène donné. Sur ces séquences consensus on sélectionne 11 sondes de 25-mers, devant être les plus spécifiques possibles. Pour chaque sonde vont ensuite être réalisés 20 oligonucléotides parfaitement homologues ou *perfect match* (PM) et 20 oligonucléotides présentant une mutation centrale ou *mismatch* (MM). Le rapport entre les taux d'hybridation des PM et des MM avec le matériel biologique permet de juger de la pertinence des résultats d'une sonde.

Longueur des sondes: 25-mer.

Format: 6,4 million de spots par puce

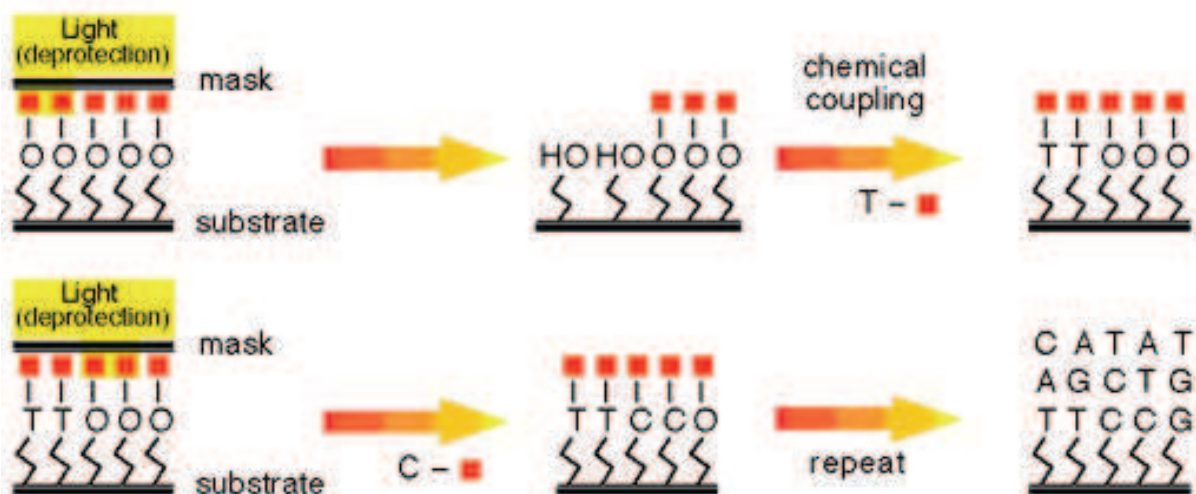


Figure 19. Schéma de synthèse *in situ* des oligonucléotides chez Affymetrix (d'après Lipshutz et al., 1999).

La lumière permet de déprotéger des régions spécifiques en ôtant le groupement photolabile.

### c) Les études pionnières du transcriptome chez *S. cerevisiae*

Les puces à ADN ont été principalement utilisées pour identifier l'ensemble des gènes dont l'expression est induite ou réprimée dans des conditions biologiques données. Sans vouloir faire une revue exhaustive de toutes les études réalisées à ce jour chez la levure, je voudrais citer ici quelques travaux qui ont contribué de façon remarquable à notre compréhension de la physiologie de *S. cerevisiae*.

Une des premières analyses de la modification globale du transcriptome chez *S. cerevisiae* a porté sur les effets de la déplétion en glucose (DeRisi et al., 1997). En présence de glucose, *S. cerevisiae* privilégie un métabolisme fermentatif, grâce auquel elle convertit

efficacement les sucres en éthanol. Lorsque le glucose est épuisé, *S. cerevisiae* peut utiliser l'éthanol produit comme substrat respiratoire. Cette transition d'un métabolisme fermentatif à un métabolisme respiratoire est aussi appelé diauxie. DeRisi et coll. ont montré que, lors de la croissance exponentielle en milieu riche, le profil global d'expression des gènes est particulièrement stable. Cependant, au fur et à mesure que la quantité de glucose dans le milieu décroît, l'expression de près de 2000 gènes est induite ou réprimée. Près de la moitié de ces gènes différentiellement exprimés n'avaient alors pas de fonction connue, et près de 400 d'entre eux ne présentaient aucune homologie avec un gène déjà caractérisé. Cette étude a ainsi pu apporter un premier indice quant à la fonction de ces gènes. Par ailleurs, les auteurs ont montré que l'étude de l'expression des gènes codant des enzymes clé du métabolisme, ou celle de larges groupes de gènes fonctionnellement reliés peut éclairer le mécanisme d'adaptation et de reprogrammation de la cellule aux changements environnementaux. Par exemple, l'expression de la plupart des gènes impliquées dans l'oxydation de l'éthanol, le cycle de Krebs et le cycle du glyoxylate est fortement induite lors d'une carence en glucose, alors que la transcription de 95% des gènes codant des protéines ribosomales est réprimée. Le travail de DeRisi et coll. a également participé à la mise en évidence des gènes dont l'expression est affectée par la délétion du répresseur transcriptionnel *TUP1* ou la sur-expression de l'activateur transcriptionnel *YAPI* ainsi qu'à l'identification de l'ensemble des gènes dépendant d'un régulateur donné. Cette étude pionnière illustre comment les variations du transcriptome peuvent apporter des informations importantes sur l'état physiologique d'une cellule et permettre l'identification de groupes de gènes co-régulés impliqués dans un processus biologique particulier. De plus, la puissance de l'utilisation des puces à ADN comme outil de caractérisation des conséquences transcriptionnelles de mutations d'un régulateur transcriptionnel donné a été clairement établie, ouvrant la voie à la notion de réseaux de régulation.

Par la suite, Spellman et coll. (Spellman et al., 1998) ont utilisés les puces à ADN pour dresser un catalogue des gènes dont le niveau d'ARNm varie périodiquement en fonction du cycle cellulaire. Après synchronisation de la croissance des cellules, la comparaison des profils d'expression a mis à jour 800 gènes régulés en fonction du cycle cellulaire dont plus de la moitié sont régulés par les cyclines *CLN3* et *CLB2*. L'examen des séquences immédiatement en amont du premier codon de ces 800 gènes a permis d'identifier des sites de fixation potentiels de facteurs, connus ou inconnus, régulant l'expression des gènes en fonction du cycle cellulaire. Cette étude a pour particularité de combiner un grand nombre d'expériences réalisées dans différentes conditions. La synchronisation des cellules a, par

exemple, été réalisée en utilisant trois méthodes différentes afin de palier à leurs artefacts caractéristiques. De plus, les auteurs ont incorporé et réanalysé les données de Cho et coll. (Cho et al., 1998). Cette impressionnante quantité de données combinée à une analyse bioinformatique rigoureuse utilisant un algorithme de Fourier afin de classer les gènes en fonction du moment de leur expression et une fonction de corrélation permettant de classer les données en clusters de gènes co-régulés est particulièrement convaincante. En plus des groupes fonctionnels déjà connus pour être régulés en fonction du cycle cellulaire (réplication et répartition de l'ADN, bourgeonnement, mitose, structure du cytosquelette, etc...), les auteurs ont identifiés d'autres groupes de gènes totalement inattendus tels que les gènes de biosynthèse de la méthionine ou les transporteurs d'hexoses. A l'issue de cette étude, le mécanisme moléculaire contrôlant le cycle cellulaire a été identifié pour à peu près 500 gènes. Cependant, pour les quelques 300 gènes restant, aucune site de fixation d'un facteur de transcription régulé en fonction du cycle cellulaire n'a pu être identifié, ouvrant ainsi de nombreuses perspectives quand à l'étude de la régulation de ces gènes. Grâce à ces données, une première carte des réseaux de contrôles transcriptionnels au cours du cycle cellulaire a pu être dessinée, et un grand nombre de nouveaux gènes ont pu être identifié comme participant à des processus cellulaires régulés au cours du cycle cellulaire.

Chu et coll. (Chu et al., 1998) ont suivi les changements dans l'expression des gènes au cours de la sporulation. Chez la levure, la sporulation englobe deux processus distincts: la méiose, au cours de laquelle une cellule diploïde va produire quatre gamètes haploïdes, ou spores, en réponse à une situation de carence sur un milieu dépourvu de composés azotés, et la morphogénèse des spores. Différentes approches, telles que des cribles génétiques, avaient jusqu'alors été utilisées pour identifier les gènes impliqués dans la sporulation. Elles démontraient que, chez la levure, la sporulation est caractérisée par la transcription séquentielle d'une centaine de gènes répartis en quatre groupes appelés précoces, moyens, moyen-tardifs, et tardifs (Mitchell, 1994). Chu et coll. ont pu identifier que la transcription de plus de 1000 gènes présente une variation significative au cours de la sporulation. Les gènes induits au cours de la sporulation ont été classés en sept profils temporels d'expression différents, reflétant la progression séquentielle de l'expression des gènes au cours de la sporulation. La classification de ces gènes a permis de déduire la fonction de gènes non caractérisés, en partant de l'hypothèse que des gènes ayant une fonction proche auront un profil d'expression similaire. Plus de la moitié des gènes identifiés sont réprimés lors de la sporulation. La transcription des protéines ribosomales est, par exemple, réprimée précocement en réponse à l'arrêt de la croissance, puis induite lors de la maturation des spores

afin de produire une quantité importante de ribosomes dans les spores. Par ailleurs, les auteurs ont montré que le facteur de transcription Ndt80 est important, sans doute en conjonction avec un autre facteur encore inconnu, pour l'induction d'une centaine de gènes à la fin de l'étape de prophase méiotique. Les auteurs ont également identifié des séquences consensus, déjà connues comme les motifs URS1 ou totalement nouvelles, potentiellement responsables de la régulation temporelle de l'expression des gènes. Ce travail montre comment les études de transcriptome peuvent élargir de manière impressionnante le nombre de gènes impliqués dans un processus biologique et apporter des indices quand à la fonction de nombreux gènes précédemment non caractérisés, dont certains ont des homologues sans doute impliqués dans la gamétogénèse chez les vertébrés.

Gasch et coll. (Gasch et al., 2000) ont utilisé les puces à ADN pour étudier les programmes d'expression géniques chez *S. cerevisiae* en réponse à des changements environnementaux. Pour ce faire, ils ont caractérisé les modifications de l'expression génique au cours du temps en réponse à un choc thermique ou osmotique, à l'exposition à des agents chimiques (peroxyde d'hydrogène, ménadione, diamine, dithiothréitol), ou encore à des carences nutritives (en acides aminés, en source d'azote, ou suite à l'entrée en phase stationnaire de croissance). Cette étude a révélé qu'une série de gènes répondent de la même façon quelle que soit la condition de stress, alors que la réponse d'autres gènes est spécifique à une condition donnée. Environ 900 gènes sont induits ou réprimés de façon similaire quelque soit la condition testée; ce phénomène a été appelé la réponse au stress environnemental ("environmental stress response" ou ESR). Les auteurs proposent que l'ESR soit une réponse générale adaptative aux environnements non-optimaux, que sa régulation dépend de différents facteurs de transcription, et implique différentes voies de signalisation qui sont à la fois gène-spécifique et condition-spécifique. La fonction des gènes caractérisés impliqués dans l'ESR suggère un mécanisme de protection de la cellule contre des conditions de stress. Ainsi, les gènes impliqués dans la synthèse de protéines ou la croissance cellulaire sont réprimés, illustrant la nécessité des cellules à conserver l'énergie. Les gènes induits au cours de l'ESR sont impliqués dans une large variété de processus allant de la génération et du stockage de l'énergie à la réparation de l'ADN, représentant des systèmes physiologiques qui doivent être protégés en toute circonstance. De plus, il a été montré que plusieurs séries de gènes co-régulés partagent des similitudes au niveau des éléments de leur promoteur, suggérant l'intervention de facteurs de transcription spécifiques à la régulation de ces gènes. L'étude de la réponse au stress de souches mutantes a suggéré l'implication des facteurs de transcription Yap1, Msn2 et Msn4, au cours de ce processus.

## **b) Distribution des protéines sur le génome: immunoprécipitation de la chromatine**

De nombreuses protéines nucléaires, impliquées dans des processus, allant de la régulation de la transcription, de la réplication ou de la réparation de l'ADN, aux modifications de structure de la chromatine, sont susceptible d'interagir avec la chromatine. Notre compréhension de la fonction de ces protéines est limitée par notre connaissance de leurs sites de liaison à la chromatine. Au cours de ma thèse, je me suis plus particulièrement intéressée à l'étude de la localisation de la machinerie de transcription sur le génome de *S. cerevisiae*, en réalisant une cartographie à grande échelle grâce à la technique de ChIP-on-chip (ou localisation à grande échelle, "genome-wide location analysis" en anglais).

La première étude de localisation à grande échelle a été réalisée en 2000 dans le laboratoire de R. Young (Ren et al., 2000) afin d'identifier les gènes dont l'expression est directement contrôlée par les activateurs transcriptionnels Gal4 et Ste12 chez *S. cerevisiae* en réponse à la présence de galactose ou de phéromones de type sexuel, respectivement. Cette méthode combine la technique d'immunoprécipitation de la chromatine ("ChIP"), qui avait déjà été utilisée pour l'étude de la localisation de protéines sur un petit nombre de sites bien définis (Solomon & Varshavsky, 1985), à une analyse par hybridation sur des puces à ADN ("on chip").

Le principe du ChIP-on-chip est décrit dans la Figure 20. Dans un premier temps, les cellules sont traitées avec un agent pontant, généralement du formaldéhyde, qui permet la fixation *in vivo* des complexes protéine-protéine et ADN-protéine par pontage covalent. Le formaldéhyde est un composé dipolaire très actif qui agit en ciblant principalement les groupements amines des acides aminés ou de l'ADN, créant ainsi des ponts entre les molécules ayant une distance d'environ 2 Å. Ces réactions se mettent en place *in vivo* en quelques minutes seulement après l'ajout de formaldéhyde sur les cellules vivantes. L'avantage de l'utilisation du formaldéhyde est que la réaction est totalement réversible. Les cellules sont ensuite lysées, et la chromatine est fragmentée par sonication afin d'obtenir des fragments d'une moyenne de 400 paires de bases. Les fragments d'ADN liés à une protéine d'intérêt sont alors enrichis par immunoprécipitation avec un anticorps spécifique. Les fragments d'ADN sont purifiés après une incubation prolongée à 65°C permettant la réversion des pontages protéine-ADN. Si la protéine d'intérêt est associée avec une région du génome spécifique *in vivo*, alors les fragments d'ADN de cette région devraient être enrichis dans l'échantillon immunoprécipité relativement aux autres régions du génome. L'efficacité de

l'immunoprécipitation peut être testée par PCR en temps réel en utilisant des amorces situées dans une région connue pour être la cible de la protéine d'intérêt. Les fragments d'ADN sont ensuite amplifiés et marqués par un fluorophore (Cy5 pour l'ADN immunoprécipité, Cy3 pour un témoin non-immunoprécipité) via une "ligation-médiée-PCR" (LM-PCR), puis hybridés de façon compétitive sur une puce à ADN. La résolution de la méthode dépend principalement de deux facteurs : la taille des fragments de chromatine soniqués et la densité des sondes présentes sur la puce.

Cette technique a rapidement été utilisée dans de nombreux laboratoires pour répondre à une large gamme de questions (Buck & Lieb, 2004; Hanlon & Lieb, 2004; Bulyk, 2006), telles que l'identification des cibles des régulateurs transcriptionnels, le rôle des modifications des histones et le remodelage de la chromatine (Barrera & Ren, 2006; Schones & Zhao, 2008; Jiang & Pugh, 2009), ou la distribution des ARN polymérases et de leurs facteurs généraux.

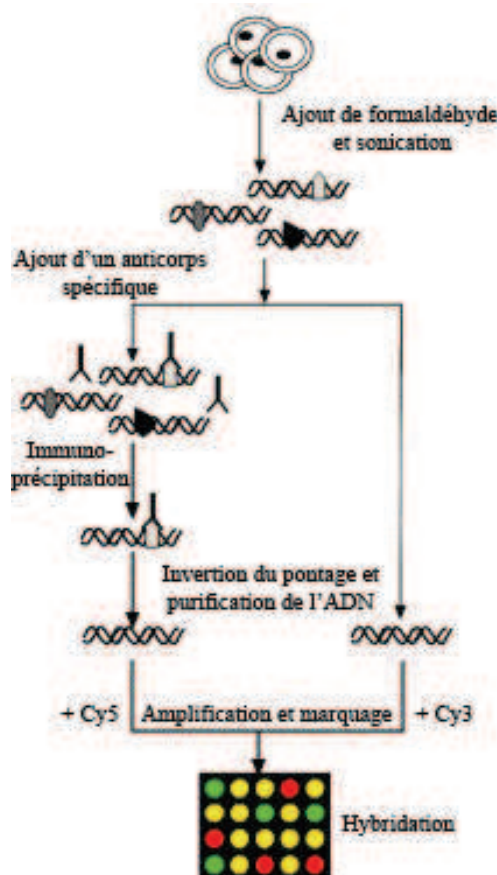


Figure 20. Résumé de la technique de ChIP-on-chip (d'après Buck & Lieb, 2004).

### Cibles des régulateurs transcriptionnels

Le profil d'expression des gènes dépend de la reconnaissance de séquences promotrices spécifiques par des régulateurs transcriptionnels. En effet, ces régulateurs permettent le recrutement et la régulation des complexes de modification de la chromatine et de la machinerie de transcription. L'identification des cibles de ces régulateurs est donc essentielle à la compréhension des réseaux de régulation complexes existants dans les cellules. Lee et coll. ont utilisé la méthode de ChIP-on-chip afin de définir les cibles de plus de 100 régulateurs transcriptionnels et de proposer une première description d'ensemble des réseaux de régulation transcriptionnels présents chez *S. cerevisiae* (Lee et al., 2002). Les auteurs ont pu identifier six motifs de réseaux de régulation différents allant de l'auto-régulation à des réseaux plus compliqués impliquant la liaison de plusieurs régulateurs différents à un même promoteur, particularité qui jusque là était réservée aux eucaryotes pluricellulaires. L'un des motifs de régulation mis à jour par cette étude est celui du "régulon ribosome". En effet, il est apparu que la protéine Fhl1, précédemment identifiée comme un suppresseur de mutations de l'ARN polymérase III (Hermann-Le Denmat et al., 1994) présente un profil inédit: elle cible quasi exclusivement tous les gènes codant des protéines ribosomales et serait ainsi un régulateur clé de la synthèse des protéines ribosomales (Schawalder et al., 2004; Wade et al., 2004; Martin et al., 2004).

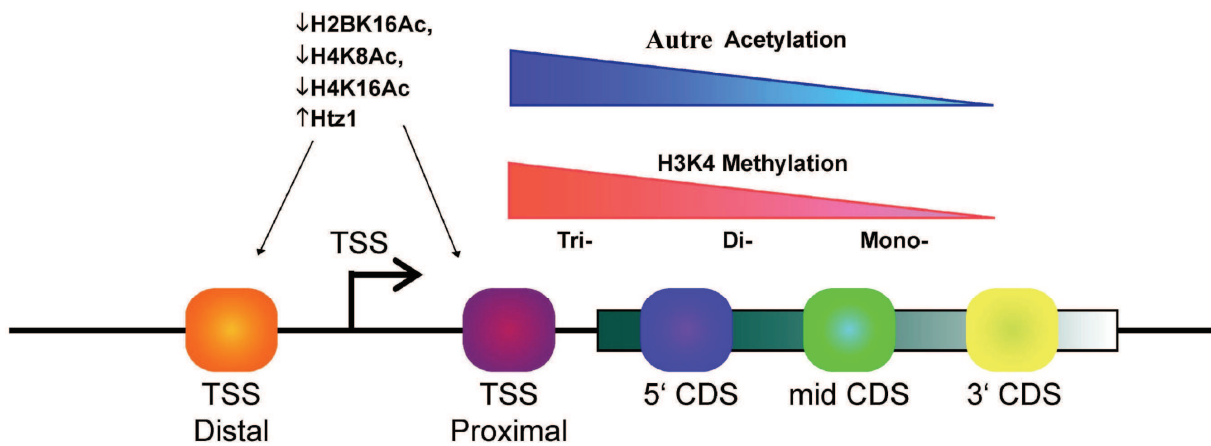
### Distribution des histones et de leurs modifications

Jusque récemment, le recrutement des histones sur l'ADN au cours de la réplication était un mécanisme mal connu. Une possibilité était que les histones soient positionnées de façon aléatoire et aient pour seule fonction d'empaqueter l'ADN. Les nucléosomes pourraient, alternativement, avoir une fonction physiologique spécifique dépendant de leur position sur le génome. Les résultats des études à grande échelle réalisées ces dernières années ont largement contribué à notre compréhension de l'importance de la structure de la chromatine dans tous les processus liés à l'ADN, et ont révélé que les cellules utilisent probablement une combinaison des deux mécanismes (dépôt au hasard et spécifique) pour l'organisation des nucléosomes. Yuan et coll. (Yuan et al., 2005) ont réalisé une cartographie de la position des nucléosomes chez la levure *S. cerevisiae*. Pour ce faire, ils ont purifié des mono-nucléosomes résultants d'une digestion du génome par la nucléase micrococcale, et ont utilisé des puces à ADN



contenant des sondes de 50 nucléotides espacées tous les 20 nucléotides le long du chromosome III ainsi que sur 223 autres régions régulatrices. Ils ont pu observer des régions d'environ 150 paires de bases dépourvues de nucléosomes situées environ 200 paires de bases en amont d'un grand nombre de régions codantes. Ces régions sont conservées au cours de l'évolution et enrichies en poly-désoxyadénine et poly-désoxythymidine. La plupart des motifs de fixation de facteurs de transcription se trouvent dans ces régions dépourvues de nucléosomes (ou NFR pour "nucleosome-free promoter region"), suggérant que la position des nucléosomes est un facteur déterminant pour l'accessibilité à la machinerie de transcription. Ces résultats concordent avec ceux de deux études préalables (Lee et al., 2004; Bernstein et al., 2004) qui montrent une déplétion en nucléosomes dans les régions promotrices des gènes activement transcrits et au niveau des motifs de fixation des facteurs de transcription. Par ailleurs, Venters et coll. ont montré que les activateurs et les remodeleurs de la chromatine sont localisés au niveau du nucléosome directement en amont de la région dépourvue de nucléosomes (appelé nucléosome -1), et que la suppression de ce nucléosome était nécessaire à l'arrivée de l'ARN polymérase II après la formation partielle du complexe de préinitiation (Venters & Pugh, 2009).

La compréhension de la distribution globale des enzymes de modification des histones et des modifications post-traductionnelles qu'elles provoquent, des variants d'histones et des remodeleurs de la chromatine est essentielle à la compréhension de la dynamique de la chromatine. Les histones sont sujettes à un grand nombre de modifications covalentes, essentiellement sur leur queue N-terminale, incluant des méthylations, acétylations, phosphorylations, ubiquitylations et des ADP-ribosylations. Celles-ci sont induites par l'action de différentes enzymes comme, par exemple, les histones acétyltransférase (HAT) ou les histones déacétylases (HDAC). Plusieurs modèles ont été proposés pour expliquer le rôle des modifications d'histones dans la transcription, le plus connu (et le plus controversé) étant celui du "code des histones" qui propose que la combinaison de multiples modifications peut dicter l'état chromatinien et la manière de réguler un gène ou un locus (Jenuwein & Allis, 2001). La fonction de ces modifications d'histones a été étudiée de manière extensive à la fois chez la levure et chez les eucaryotes pluricellulaires. Il a par exemple été montré, chez la levure, que la méthylation de la lysine 4 de l'histone H3 (H3K4) est corrélée positivement avec le taux de transcription (Pokholok et al., 2005) et que le degré de méthylation (mono-, di-, tri-) diminue de la région 5' vers la région 3' des gènes (Liu et al., 2005) (Figure 21).



**Figure 21.** Vue schématique de l'architecture de la chromatine chez *S. cerevisiae* (d'après Liu et al., 2005).

Les gènes de *S. cerevisiae* sont généralement caractérisés par une région dépourvue de nucléosomes qui sert de site de démarrage de la transcription ("Transcription Start Site", TSS). Les deux nucléosomes entourant cette région ont de faibles niveaux d'acétylation des histones H2BK16, H4K8 et H4K16 et utilisent Htz1 à la place de l'histone canonique H2A. Les autres acétylations diminuent selon un gradient 5' vers 3' le long des gènes activement transcrits. De plus, les gènes activement transcrits présentent un gradient de méthylation de l'histone H3K4, avec une majorité de tri-méthylations en 5', puis des di- et mono-méthylations à mesure que l'on s'approche de l'extrémité 3'.

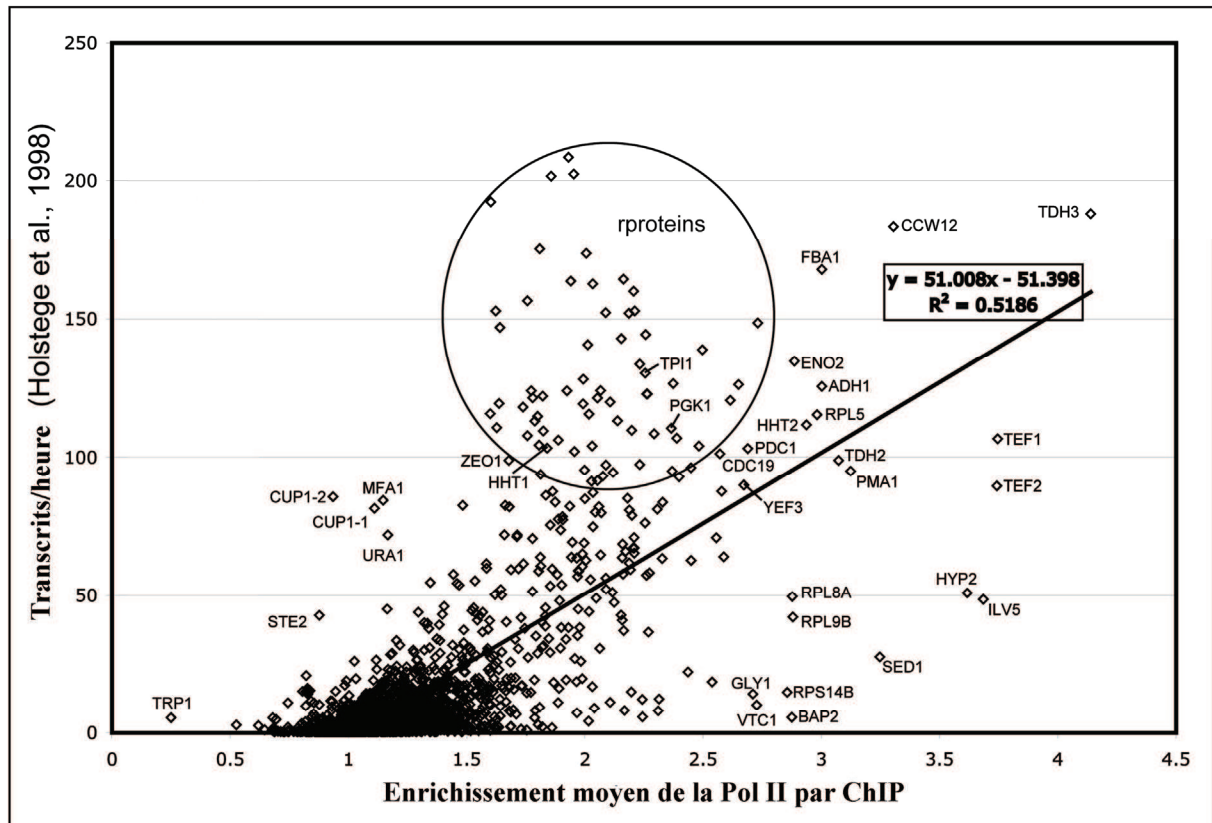
### Distribution des ARN polymérases et de leurs facteurs généraux

Un autre domaine dans lequel la cartographie à grande échelle a été d'une importance prépondérante est l'étude de la machinerie générale de transcription. L'une des premières protéines dont la distribution ait été inventoriée de façon systématique est l'ARN polymérase III de *S. cerevisiae* (Harismendy et al., 2003; Roberts et al., 2003; Moqtaderi & Struhl, 2004). La localisation de la machinerie de transcription de l'ARN polymérase III a été déterminée en analysant en parallèle la distribution de l'ARN polymérase III et de ses deux facteurs généraux de transcription TFIIIB et TFIIIC. Comme attendu, ces études ont permis de confirmer la présence des transcrits déjà connus de l'ARN polymérase III (ARNs de transfert et quelques autres ARN non-codants). Elles ont pu également mettre à jour des gènes jusqu'alors inconnus

tels que *SNR52*, un snoARN dont la transcription avait précédemment été imputée à l'ARN polymérase II, et *ZOD1* (pour "zone of disparity") dont la fonction est inconnue. De plus, huit régions présentent une occupation significative par TFIIC mais pas par TFIIB ou l'ARN polymérase III; ces régions ont été appelées *ETC* (pour "extra TFIIC"). Par ailleurs, ces études ont montré que l'expression des gènes de classe III est modulée au cours d'une carence en nutriment ou en phase stationnaire de croissance par une diminution du recrutement de l'ARN polymérase III et de TFIIB, alors que TFIIC reste lié à ses séquences cibles.

Des études plus récentes ont été consacrées à la localisation de l'ARN polymérase II sur l'ensemble du génome de la levure. Un premier travail a comparé sa distribution dans une souche sauvage à celle d'un mutant de l'hélicase Sen1 (Steinmetz et al., 2006). Sen1 est un facteur de terminaison de l'ARN polymérase II à priori spécifique des ARN non-codants mais qui pourrait également avoir un rôle plus général dans la terminaison de la transcription par l'ARN polymérase II, et dont l'homologue humain est impliqué dans des maladies neuro-dégénératives. Les auteurs montrent qu'il est possible de détecter les gènes transcrits par l'ARN polymérase II en examinant le profil d'occupation de cette protéine sur le génome, et ainsi identifier de nouvelles régions transcrites par l'ARN polymérase II. Une corrélation de ces résultats avec des données de transcriptome (Holstege et al., 1998) montre une relation relativement linéaire entre ces deux jeux de données (Figure 22). Cependant, une large proportion de protéines ribosomales présente un profil différent, avec un enrichissement relativement modéré par l'ARN polymérase II mais un taux de transcription très élevé. La raison de ce phénomène est inconnue, mais elle pourrait indiquer un mode de transcription spécifique de ces gènes ou une différence entre les conditions de cultures. Par ailleurs, les auteurs ont pu confirmer que les snARN sont la cible principale de Sen1. Ils ont également montré qu'un certain nombre d'autres gènes ont une terminaison Sen1-dépendante. Ces gènes sont en général de petite taille, suggérant que le mode d'action de Sen1 permet une reconnaissance préférentielle des éléments terminateurs se trouvant à proximité de l'extrémité 5' du gène.

Les expériences à grande échelle ont commencé à modifier l'idée selon laquelle l'étape limitante de la transcription de la plupart des gènes est le recrutement de l'ARN polymérase II. En effet, il a été montré qu'en phase stationnaire de croissance, l'ARN polymérase II est localisée au promoteur de plusieurs centaines de gènes dont la transcription est réprimée dans ces conditions mais qui sont nécessaires à une réponse rapide en sortie de phase stationnaire



**Figure 22. Comparaison du nombre de transcrits par heure au ratio d'enrichissement moyen de l'ARN polymérase II (d'après Steinmetz et al., 2006).**

Une régression linéaire et son équation sont représentées. Quelques points sélectionnés sont représentés avec le nom de l'ORF. Le cercle définit 61 gènes codant des protéines ribosomales et trois autres gènes pour lesquels l'enrichissement de l'ARN polymérase II est inhabituellement faible par rapport à leur taux de transcription.

chez *S. cerevisiae* (Radonjic et al., 2005). Par ailleurs, Venters et coll. ont montré que plus de la moitié de gènes transcrits par l'ARN polymérase II chez *S. cerevisiae* sont faiblement transcrits et ne présentent que peu d'ARN polymérase II localisée dans leur région codante (Venters & Pugh, 2009). La présence d'ARN polymérase peut cependant être détectée dans la région promotrice ou au début de la région codante de ces gènes, suggérant ainsi que la transcription de la plupart des gènes est régulée à la fois au niveau du recrutement, mais aussi après le recrutement de l'ARN polymérase II. L'existence de gènes pour lesquels l'ARN polymérase II se trouve en état de pause à proximité du promoteur (en anglais "promoteur proximal pausing", i.e. présence de l'ARN polymérase au niveau du promoteur mais pas de signal ou signal faible dans la région codante) est connue depuis de nombreuses années. On sait depuis assez longtemps que, chez *Drosophila melanogaster*, l'ARN polymérase II est significativement enrichie au niveau de la région promotrice du gène codant la protéine de choc thermique HSP70 avant et après induction des gènes de choc thermique (Gilmour & Lis, 1985, 1986). Plusieurs études récentes réalisées chez les eucaryotes pluricellulaires, et plus particulièrement la drosophile, montrent que ce phénomène est bien plus étendu (Muse et al., 2007; Zeitlinger et al., 2007), puisque l'ARN polymérase II est présente en situation de pause sur le promoteur de 10 à 15% des gènes, permettant ainsi une réponse rapide à des stimuli ou à des signaux développementaux. Des résultats similaires ont également été obtenus dans des cellules souches embryonnaires humaines (Guenther et al., 2007).

En plus de l'ARN polymérase II, plusieurs études ont été réalisées sur la localisation globale de ses facteurs de transcriptions. La localisation du Médiateur a, par exemple, été examinée via différentes sous-unités des quatre modules composants ce complexe chez *S. cerevisiae* (Andrau et al., 2006) et *S. pombe* (Zhu et al., 2006). Comme attendu, le Médiateur est localisé dans les régions promotrices des gènes activement transcrits. Cependant, il a également été retrouvé dans les régions promotrices de gènes inactifs et dans la région codante de quelques gènes activement transcrits. De plus, le module répresseur Cdk8 semble associé au Médiateur sur tous ses sites de liaison, mais présente un degré d'occupation réduit, suggérant qu'il pourrait interagir avec le Médiateur de façon transitoire. La comparaison de ces résultats avec ceux de la localisation de l'ARN polymérase II en phase stationnaire de croissance permet d'émettre l'hypothèse selon laquelle le Médiateur jouerait un rôle de plateforme pour l'ARN polymérase II au promoteur des gènes inactifs durant la quiescence.

Bien que le ChIP-on-chip soit une technique particulièrement puissante pour les études de génomique fonctionnelle, elle possède un certain nombre de limitations notables. La

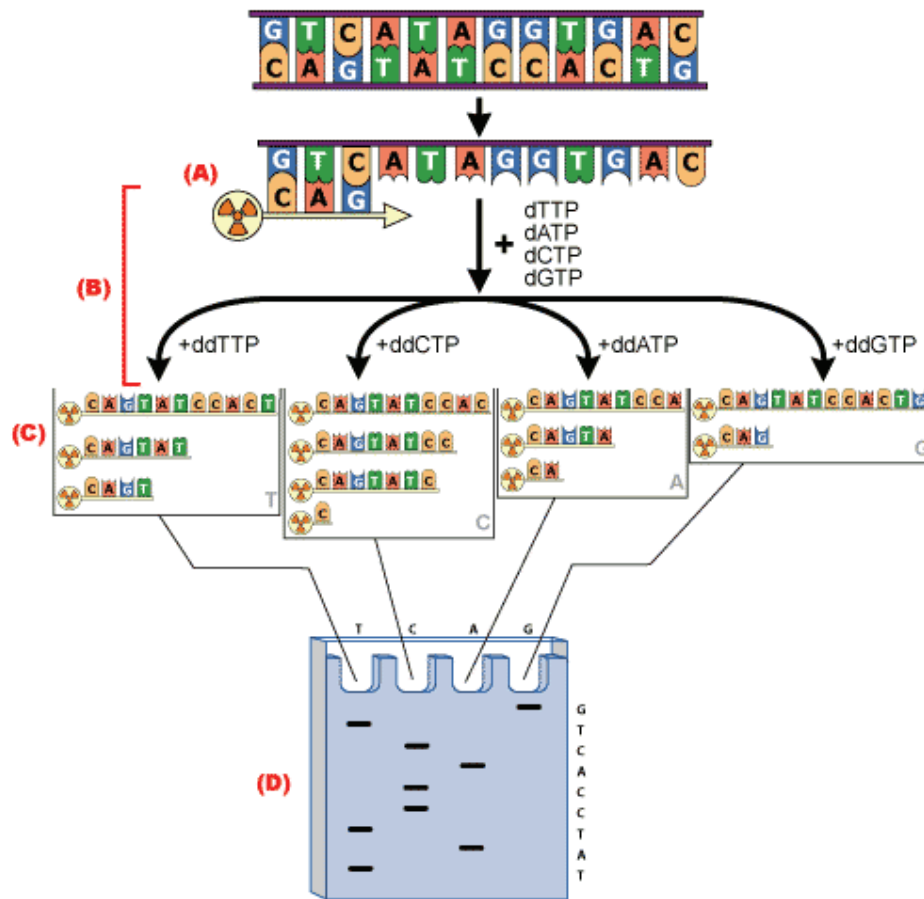
première limitation est (malheureusement) le prix des puces à ADN et de tout le matériel nécessaire qui reste prohibitif pour un certain nombre de laboratoires. Chaque expérience doit être répétée, au minimum, trois fois pour permettre une analyse statistique robuste, sachant qu'une expérience ne représente la distribution que d'une seule protéine dans une condition donnée. D'un point de vue technique, l'analyse des données issues des expériences de ChIP-on-chip reste difficile car elle nécessite l'utilisation de procédés de normalisation complexes pour extraire la réalité biologique des artefacts techniques. La communauté scientifique s'est abondamment penché sur la question au cours de la dernière décennie, aboutissant au développement d'outils tels BioConductor (Gentleman et al., 2004), mais le recours à des bio-informaticiens compétents reste souvent de mise. La résolution obtenue dans les expériences de ChIP-on-chip a considérablement augmenté. Tandis que les premières études utilisaient des puces à ADN portant une sonde par ORF et par région intergénique, aujourd'hui les puces à ADN couvrent l'ensemble d'un génome avec des sondes qui peuvent même être chevauchantes ("tilling array"), permettent potentiellement d'obtenir une résolution au nucléotide près. Cependant, la taille des fragments d'ADN obtenus après sonication peut difficilement être inférieure à 100 paires de bases, limitant ainsi la résolution obtenue. L'étape de pontage au formaldéhyde peut parfois être limitante. En effet, la nature exacte des liaisons formées par le formaldéhyde sur la chromatine *in vivo* n'est pas bien caractérisée, et l'efficacité de pontage peut varier d'une protéine à l'autre. Une autre limitation importante est la disponibilité d'anticorps de qualité utilisés pour l'immunoprécipitation de la chromatine, qui doivent être à la fois spécifiques et pouvoir reconnaître un épitope qui soit accessible. Une possibilité pour surmonter cette limitation est la fusion de la protéine d'intérêt à une étiquette, par exemple de type HA ou Flag, pour laquelle des anticorps de qualité sont disponibles. L'étiquetage est un procédé largement utilisé chez la levure de part l'efficacité de la recombinaison homologue chez cet organisme, mais peut s'avérer plus difficile chez les eucaryotes pluricellulaires.

### **c) Apports récents du séquençage à haut-débit**

Le séquençage de l'ADN a été inventé à la fin des années 1970, de façon indépendante par l'équipe de Walter Gilbert (Maxam & Gilbert, 1977) et celle de Frederick Sanger (Sanger & Coulson, 1975; Sanger et al., 1977). Ces deux méthodes sont fondées sur des principes

diamétralement opposés: l'approche de Sanger est une méthode par synthèse enzymatique sélective, tandis que celle de Maxam et Gilbert est une méthode par dégradation chimique sélective. Au cours des 25 dernières années, la méthode de Sanger a été largement développée grâce à plusieurs avancées technologiques importantes et a progressivement supplanté la méthode de Maxam et Gilbert. Le principe en lui-même a cependant peu changé. Il repose sur la synthèse d'un brin d'ADN complémentaire par une ADN polymérase en présence de désoxynucléotides (dNTP) et de didésoxynucléotides (ddNTP) permettant une terminaison irréversible de la synthèse. La polymérisation d'ADN est interrompue au hasard chaque fois qu'un ddNTP est ajouté au brin en cours de synthèse. Les différents produits obtenus sont séparés sur un gel de polyacrylamide. Un traceur fluorescent (ou radioactif) permet de détecter les fragments d'ADN synthétisés (Figure 23).

Depuis leur apparition en 2005, les technologies de séquençage à haut débit ("high-throughput sequencing" ou "next generation sequencing" en anglais) ont eu un impact considérable sur la recherche en génomique. Elles ont des applications très variées allant du séquençage ou re-séquençage standard des génomes à la métagénomique en passant bien sûr par la génomique fonctionnelle. La particularité de ces technologies est leur coût réduit et la possibilité de produire un très grand nombre de séquence en un temps très court. Les 3 principales plateformes de séquençage à haut débit utilisées à ce jour sont le 454 commercialisé par Roche, le Solexa (Illumina) et le SOLiD (Applied Biosystems) (Morozova & Marra, 2008). Leurs principales caractéristiques sont résumées ci-dessous (Figure 24):



**Figure 23. Séquençage par la méthode de Sanger.**

**A.** L'ADN double brin est dissocié et une sonde radioactive ou fluorescente y est hybridée. **B.** Quatre réactions séparées sont réalisées afin de synthétiser un nouvel ADN à partir de la sonde. Chaque réaction contient les quatre désoxynucléotides et une petite quantité de l'un des didésoxynucléotides. **C.** L'ADN de chacune des quatre réactions est séparé sur un gel de polyacrylamide. **D.** La séquence peut être lue sur le gel.



- 454 : pyroséquençage à haut débit

Le séquençage par la technologie 454 se déroule en quatre principales étapes. Dans un premier temps, l'ADN génomique est fragmenté par nébulisation, deux adaptateurs sont fixés par ligation aux deux extrémités. Dans un deuxième temps, les molécules d'ADN simple brin sont fixées à des microbilles possédant à leur surface des amorces complémentaires à un des adaptateurs. Chaque microbille porte un fragment unique d'ADN simple brin. Les microbilles porteuses des brins d'ADN sont mises en émulsion en présence des réactifs pour PCR. Chaque goutte (microréacteurs) englobe une microbille et donc une molécule d'ADN, ce qui permet une amplification clonale de chaque fragment. Après amplification les microgouttelettes sont dissociées, et les microbilles porteuses de l'ADN amplifié sont transférées dans une plaque de picotitration. Dans un troisième temps, le séquençage est réalisé selon le principe de pyroséquençage qui mesure le relargage de phosphate inorganique (PPi) par chimioluminescence. Il s'agit d'un séquençage par synthèse ("sequencing-by-synthesis") qui se caractérise par la révélation en temps réel de l'activité de l'ADN polymérase qui ajoute un seul nucléotide non fluorescent à la fois. L'ADN est immobilisé et les solutions de dNTP sont ajoutées en une seule fois. Chaque fois qu'un nucléotide est ajouté, le relargage du PPi peut être détecté par une émission de lumière grâce à la présence d'une enzyme chimioluminescente. Une caméra permet de capturer les images après addition de chaque nucléotide, puis reproduit sous forme d'un pic sur un "pyrogramme". La hauteur de ce pic est fonction de l'intensité du signal lumineux, elle-même proportionnelle au nombre de nucléotides incorporés en même temps. On peut donc déduire la séquence à partir de la taille des pics obtenus. Pour finir, les données obtenues sont analysées en utilisant différents outils bioinformatiques. La dernière génération d'appareils 454 permet de générer des séquences de 400 paires de base, et plus d'un million de paires de bases en 10 heures.

- Solexa : séquençage par synthèse via des terminateurs réversibles

Le principe du séquençage par la technologie Solexa est basé sur l'incorporation réversible de nucléotides fluorescents (CRT : "cyclic reversible termination") et la lecture optique de la fluorescence. L'ADN génomique est fragmenté par nébulisation et des adaptateurs sont fixés sur chaque extrémité par ligation. L'ADN est alors fixé sur des cellules ("flow cell") grâce aux adaptateurs. Une étape d'amplification permet l'obtention de "clusters" composés de plusieurs milliers de copies de chaque fragment d'ADN. Ceux-ci sont séquencés de façon massivement parallèle grâce à une approche de séquence par synthèse qui utilise des terminateurs marqués par des fluorophores de quatre couleurs différentes. La séquence de

chaque "cluster" est obtenue par lecture de la couleur après l'addition successive de chaque nucléotide. La technologie Solexa permet de générer des séquences de 25-30 paires de base, et plus d'un milliard de paires de bases en 2-3 jours.

- SOLiD: séquençage par hybridation/ligation

La technologie SOLiD permet de réaliser un séquençage massivement parallèle par hybridation/ligation. Après fragmentation de l'ADN génomique et ajout d'adaptateur, une étape de PCR similaire à celle du 454 est réalisée. Les produits d'amplification sont alors transférés sur une surface vitrée où le séquençage est réalisé par des séries successives d'hybridations et de ligations avec 16 combinaisons différentes de dinucléotides marqués par 4 différents fluorophores. L'identité de chaque nucléotide est déterminée par l'analyse de la couleur qui résulte de deux réactions de ligation successives. La technologie SOLiD permet de générer des séquences de 35 paires de bases, et plus d'un milliard de paires de bases en 8 jours.

Les techniques de séquençage ont apporté un nouveau souffle à l'étude de la distribution des protéines sur la chromatine. Elles ont en effet conduit à l'émergence de la technique de "ChIP-Seq", qui repose sur l'immunoprécipitation de la chromatine suivie du séquençage à haut débit des fragments d'ADN obtenus. Le ChIP-Seq présente un certain nombre d'avantages comparé au ChIP-on-chip: un moindre coût, la nécessité de réaliser moins de réplicats, l'utilisation de moins de matériel biologique de départ. La recherche des sites de fixation se fait sans à-priori puisqu'elle ne nécessite pas l'utilisation de sondes prédéfinies. Cette technique est particulièrement adaptée aux grands génomes qui, malgré les avancées en termes de miniaturisation des puces à ADN, ne tiennent pas encore dans leur intégralité sur une seule puce. De plus, l'analyse des résultats est relativement plus aisée car elle ne requiert pour ainsi dire aucune étape de normalisation, bien que la quantité considérable de données générées nécessite une capacité de stockage et de traitement importante. Les premiers résultats obtenus grâce à cette technique ont permis de générer, à très haute résolution, une carte de la distribution génomique de 20 méthylations différentes des lysines et arginines des histones, ainsi que de la localisation du variant d'histone H2A.Z, de l'ARN polymérase II et de la protéine CTCF chez l'homme (Barski et al., 2007). Par la suite, cette technique a été utilisée pour étudier, entre autres, la localisation de différentes protéines sur la chromatine

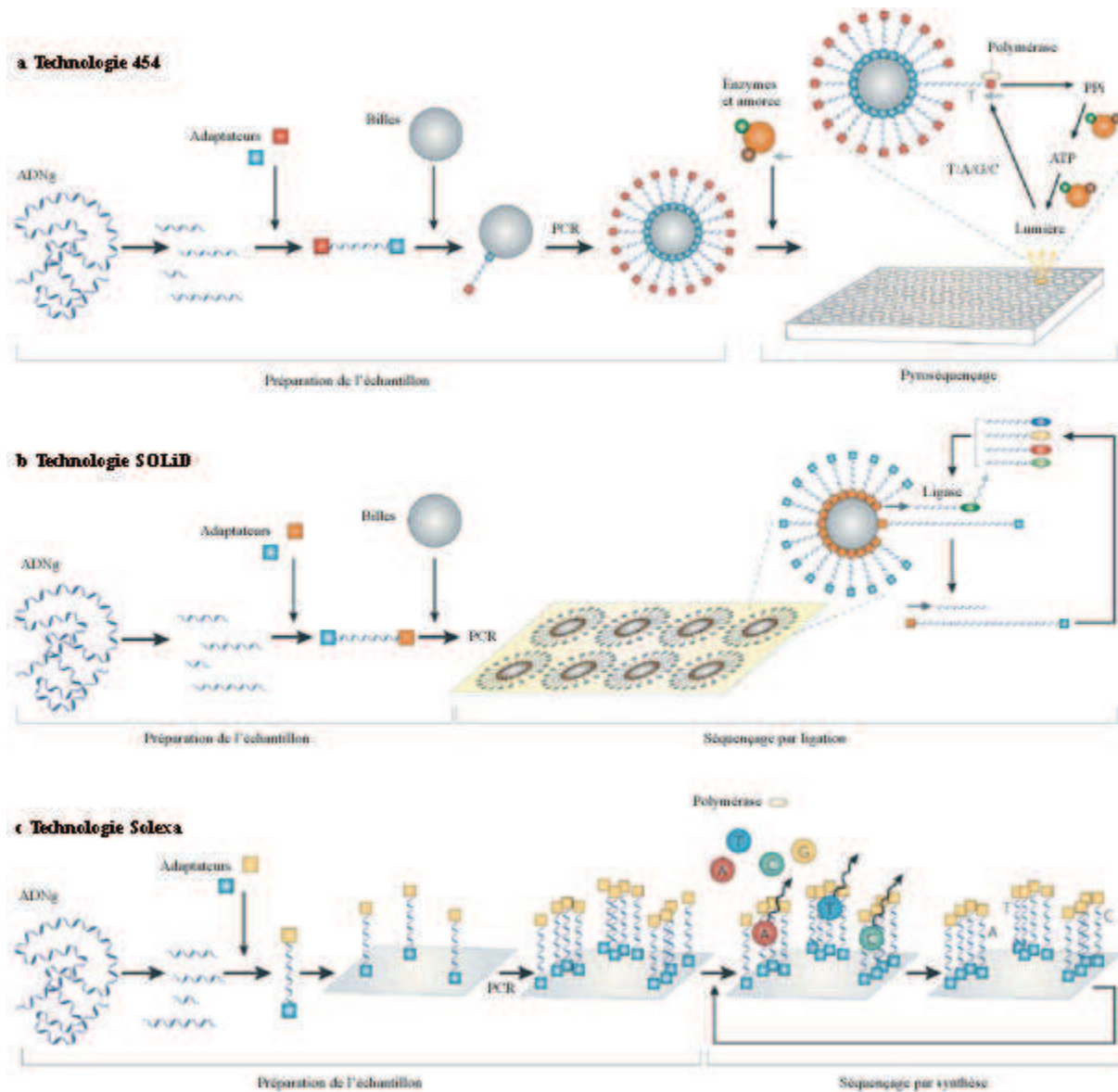


Figure 24. Les technologies de séquençage à haut débit (d'après Medini et al., 2008).

A. La méthode de séquençage 454 repose sur une technique de pyroséquençage. B. La technologie SOLiD utilise une hybridation et une ligation de dinucléotides marqués. C. La technologie Solexa utilise des terminateurs fluorescents et un séquençage par synthèse.

(Robertson et al., 2007), les modifications d'histones, la position des nucléosomes, la méthylation de l'ADN, etc... (Schones & Zhao, 2008).

L'un des derniers domaines dans lesquels le séquençage à haut débit a démontré son extrême efficacité est l'identification des ARN non-codants. La technique de "RNA-seq", développée chez *S. cerevisiae*, passe par le séquençage des ADNc obtenus à partir de l'ensemble des ARNs de la cellule (Nagalakshmi et al., 2008). L'étude du transcriptome de la levure par RNA-seq a, entre autres, permis de dresser un catalogue très précis de toutes les régions transcrites, et a montré qu'une proportion très importante (75%) du génome est activement transcrite. Plusieurs études réalisées chez différents organismes en utilisant le séquençage à haut débit ou des puces à ADN ont mis à jour une nouvelle classe de transcrits dont le site d'initiation est très proche du site d'initiation de la transcription d'une séquence codante (Preker et al., 2008; Seila et al., 2008; He et al., 2008; Core et al., 2008). Ces ARN sont courts, présents en très faible abondance, et sont souvent transcrits dans le sens opposé de la région codante correspondante, faisant ainsi apparaître la possibilité que la majeure partie des promoteurs soient intrinsèquement bidirectionnels.

Différentes hypothèses ont été émises pour expliquer la présence de ces transcrits. L'une d'entre elles serait qu'il s'agit de régions transcrites uniquement parce qu'elles sont situées dans une zone dépourvue de nucléosomes, qui en temps normal préviennent l'accessibilité des sites d'initiation de la transcription potentiels. Il est possible que ces transcrits résultent d'une ARN polymérase "pausée", qui serait incapable de poursuivre l'élongation et se décrocherait juste après l'initiation. Il semble cependant peu probable qu'un si grand nombre de transcrits soit produits sans qu'ils aient une quelconque fonction. Des études récentes chez la levure *S. cerevisiae* ont en effet montré que certains ARNs non-codants peuvent avoir un rôle régulateur. Dans le cas du gène *SER3*, l'initiation de la transcription à un site présent en aval masque le promoteur de *SER3* et réprime ainsi sa transcription (Martens et al., 2004). La présence de transcrits instables, ou "CUTs" (en anglais "Cryptic Unstable Transcripts"), normalement dégradés par l'exosome, avait été précédemment identifiée chez *S. cerevisiae* (Wyers et al., 2005). Deux études récentes ont montré que ces CUTs sont très largement répandus dans le génome, et ont également suggéré le caractère bidirectionnel des promoteurs chez *S. cerevisiae* (Neil et al., 2009; Xu et al., 2009). Bien que certains CUTs sont clairement impliqués dans le contrôle de l'expression des gènes (Camblong et al., 2007), le rôle global de tous ces ARNs non-codants reste encore mystérieux. Constituent-ils un "bruit" transcriptionnel, ou sont-ils tous impliqués dans la régulation de l'expression génique? Cette question reste encore ouverte à l'heure actuelle et

sera illustrée dans la partie Résultats à travers l'étude de la transcription du gène *URA2* chez la levure.

## **RESULTATS ET DISCUSSION**



## ***I/ Etude de la localisation globale du facteur TFIIS***

### **a) Rôle de TFIIS dans la transcription par l'ARN polymérase III chez *S. cerevisiae***

#### Contexte du travail

Trois ARN polymérases se partagent la transcription des génomes eucaryotes. Des complexes multi-protéiques de grande taille, organisés autour de l'ARN polymérase II, de facteurs généraux de transcription, de coactivateurs transcriptionnels tels le Médiateur et de facteurs d'élongation participent à la transcription des gènes de classe II (transcrits par l'ARN polymérase II). L'étude de ces complexes a été renouvelée par l'utilisation de techniques globales d'immunoprécipitation des protéines liées à la chromatine, permettant de suivre l'occupation des complexes de transcription sur l'ensemble du génome.

La protéine TFIIS est un facteur général de transcription très conservé, dont l'étude intéresse particulièrement le laboratoire en raison de ses interactions fonctionnelles avec le Médiateur (Wery et al., 2004; Guglielmi et al., 2007). TFIIS est connu depuis de nombreuses années comme étant un facteur d'élongation de l'ARN polymérase II, requis pour la stimulation de son activité de clivage intrinsèque lorsque celle-ci rencontre des sites de pause au cours de la transcription (Fish & Kane, 2002). Ces dernières années, plusieurs équipes ont montré que TFIIS joue également, en conjonction avec le Médiateur, un rôle dans l'initiation de la transcription par l'ARN polymérase II (Malagon et al., 2004; Prather et al., 2005; Guglielmi et al., 2007; Kim et al., 2007).

Bien que le rôle de TFIIS dans la stimulation de l'activité de clivage de l'ARN polymérase II soit relativement bien décrit *in vitro*, celui-ci reste moins clair *in vivo*. Il n'est par exemple pas clairement établi si TFIIS est constamment associé à l'ARN polymérase II ou bien s'il est spécifiquement recruté lorsque l'ARN polymérase II rencontre un site de pause. De plus, le rôle récemment découvert de TFIIS dans l'initiation de la transcription par l'ARN polymérase II suggère que cette protéine pourrait avoir d'autres rôles encore inconnus. Nous avons donc voulu étudier la localisation globale de TFIIS sur le génome de la levure *S. cerevisiae* par des expériences de ChIP-on-chip, afin de mieux comprendre la fonction de cette protéine *in vivo*. Pour ce faire, nous avons utilisé des puces à ADN commerciales (Agilent "Yeast Whole Genome ChIP-on-chip Microarrays" de type G4486A ou G4493A)



dont l'utilisation a dû au préalable être mise en place au laboratoire. Ces puces sont composées de 41418 (G4486A) ou 41776 (G4493A) sondes réparties sur environ 85% de la portion non répétitive du génome de *S. cerevisiae*, avec une résolution moyenne de 250 à 300 nucléotides.

L'étude de la localisation globale de TFIIS sur le génome de *S. cerevisiae* et sa comparaison avec celle de l'ARN polymérase II révèle une distribution bimodale. Ainsi, sur les gènes de classe II, nous observons une bonne corrélation entre ces deux jeux de données. Une deuxième catégorie de sondes, majoritairement situées dans des régions intergéniques sont spécifiquement liées par TFIIS mais pas par l'ARN polymérase II; leur analyse révèle qu'il s'agit de gènes de classe III (voir ci-dessous). TFIIS est présent sur près de la moitié des gènes de classe II, mais il est à noter que le signal de TFIIS étant par nature plus faible que celui de l'ARN polymérase II, sa présence n'est pas toujours détectable sur les régions faiblement liées par l'ARN polymérase II. La présence de TFIIS a également été détectée sur des régions intergéniques dont la transcription produit des ARNs non-codants, tels que *NEL025C* (Wyers et al., 2005) ou *SRGI* (Martens et al., 2004). L'observation de l'occupation de TFIIS sur plusieurs gènes fortement transcrits par l'ARN polymérase II révèle un enrichissement constant tout au long de l'ORF dont le profil suit précisément celui de l'ARN polymérase II. Ce résultat va à l'encontre d'un précédent modèle selon lequel TFIIS ne serait recruté qu'en présence de MPA, condition mimant une carence en nucléotides dans la cellule (Pokholok et al., 2002), mais confirme des résultats plus récents obtenus sur un petit nombre de gènes (Prather et al., 2005; Guglielmi et al., 2007). La comparaison de l'occupation globale de TFIIS sur le génome en présence ou en absence de MPA, dont l'ajout dans le milieu de culture affecte l'élongation de la transcription en diminuant la quantité de nucléotides dans la cellule, n'a par ailleurs révélé qu'une faible augmentation de la présence de TFIIS dans ces conditions.

L'analyse de l'occupation globale de TFIIS sur le génome de *S. cerevisiae* par ChIP-on-ChIP a également révélé, de façon inattendue, la présence de cette protéine au niveau de tous les gènes de classe III (*ADNr*, *SCR1*, *RPRI*, *SNR52*, *SNR6*, *ZOD1*) à l'exception de l'ARNr 5S, suggérant ainsi un rôle jusqu'alors inconnu de TFIIS dans la transcription par l'ARN polymérase III. Aucun enrichissement n'a pu être détecté dans le cas des régions *ETC*, occupée par TFIIC mais pas par TFIIB ni par l'ARN polymérase III (Moqtaderi & Struhl, 2004). Dans le cas du gène *SNR6*, le pic d'occupation de TFIIS et de l'ARN polymérase III semble être étrangement localisé au niveau de la boîte B. La comparaison de l'occupation de TFIIS avec celle de l'ARN polymérase III sur les gènes de classe III révèle une excellente

corrélation entre les deux jeux de données. Cette corrélation a également été confirmée par des expériences d'immunoprécipitation de la chromatine le long du gène *SCR1*, le seul gène suffisamment long pour étudier la distribution de la machinerie de transcription par l'ARN polymérase III et de TFIIS. Par ailleurs, nous avons pu déterminer que le rôle de TFIIS dans la transcription par l'ARN polymérase III est indépendant de son rôle dans la transcription par l'ARN polymérase II. Pour ce faire, nous avons analysé l'occupation par TFIIS de plusieurs gènes de classe II et III lorsque la transcription par l'ARN polymérase II ou III est spécifiquement affectée dans des souches mutantes pour les sous-unités de l'une ou l'autre des ARN polymérases.

Le rôle de TFIIS dans la transcription par l'ARN polymérase II requiert une interaction directe entre les deux protéines *via* le domaine II. Nous avons donc réalisé des expériences de co-immunoprécipitation afin d'explorer une possible interaction entre TFIIS et l'ARN polymérase III. Nous n'avons pas pu détecter d'interaction entre TFIIS et l'ARN polymérase III ou les facteurs TFIIB ou TFIIC à partir d'extraits totaux. Cependant, la même expérience réalisée à partir d'extraits obtenus après pontage au formaldéhyde révèle que TFIIS co-immunoprécipite avec l'ARN polymérase III et TFIIC, mais pas TFIIB, démontrant la présence concomitante de ces protéines sur les gènes de classe III.

L'ARN polymérase III possédant une activité de clivage intrinsèque, dépendante de sa sous-unité Rpc11, nous avons dans un premier temps été surpris par la présence de TFIIS sur les gènes de classe III. Nos résultats suggèrent que ces deux protéines jouent un rôle important mais complémentaire dans la transcription par l'ARN polymérase III. En effet, TFIIS n'est pas capable de compléter l'effet de la délétion de la sous-unité Rpc11, aussi bien *in vivo* (la sous-unité Rpc11 est une sous-unité essentielle) que *in vitro* (l'ajout de TFIIS dans un système de transcription par l'ARN polymérase III dépourvue de Rpc11 ne restaure pas l'activité de clivage de cette enzyme). De plus, des expériences d'immunoprécipitation de chromatine et de co-immunoprécipitation après pontage au formaldéhyde suggèrent que TFIIS et Rpc11 sont présentes de façon simultanée sur les gènes de classe III.

Des expériences complémentaires ont montré que TFIIS se comporte, *in vivo*, comme un facteur général de la transcription par l'ARN polymérase III. Nous avons ainsi pu montrer que la délétion de *DST1*, le gène codant le facteur TFIIS chez *S. cerevisiae*, diminue la capacité de la cellule à supprimer la mutation *ade2-1* (le suppresseur de cette mutation, *SUP11*, étant transcrit par l'ARN polymérase III). Par ailleurs, des expériences de Northern Blot et de transcriptome ont révélé que la délétion de *DST1* réduit le niveau de transcription d'un certain nombre de gènes de classe III à 16°C (condition affectant la croissance d'un

mutant *dst1Δ*). Dans un mutant *dst1Δ*, nous avons pu observer une diminution globale de l'occupation des ARN polymérases II et III sur le génome à 16°C, bien que dans le cas de l'ARN polymérase II des effets indirects puissent expliquer une plus grande variabilité dans les conséquences de cette mutation. L'étude de deux mutants ponctuels, le mutant *dst1-E291A* dans le domaine RSADE de TFIIS et le mutant *dst1-R200A* dans le domaine de liaison à l'ARN polymérase II de TFIIS, a révélé que le rôle de TFIIS dans la transcription par l'ARN polymérase III serait probablement dépendante de son domaine RSADE. En effet, alors que ces deux mutations affectent l'occupation de l'ARN polymérase II à 16°C, seul la mutation *dst1-E291A* réduit de façon globale et importante l'occupation de l'ARN polymérase III sur ses cibles. Des expériences d'immunoprécipitation de la chromatine classique ont par ailleurs révélé que la délétion de *DST1* affecte également la présence du facteur TFIIB, mais pas de TFIIC, sur plusieurs gènes de classe III.

Afin d'étudier l'influence de TFIIS sur la transcription par l'ARN polymérase III *in vitro*, nous avons cloné le gène *DST1* sauvage ainsi que le gène contenant la mutation *dst1-E291A* dans un vecteur d'expression chez *E. coli*, produit les deux protéines, et les avons purifiées à homogénéité. Ces deux protéines ont ensuite pu être utilisées dans un système de transcription *in vitro* contenant l'ARN polymérase III purifiée et les facteurs TFIIC et TFIIB recombinants (Ducrot et al., 2006). L'utilisation de ce système pour la transcription du gène *SUP4 in vitro* génère, en plus du transcrit attendu, plusieurs transcrits plus courts migrant sous la forme d'une ou deux bandes supplémentaires sur un gel de polyacrylamide. Nous avons pu déterminer que ces transcrits sont issus d'évènement d'initiation incorrect. En effet, ils disparaissent lorsque la réaction de transcription est réalisée en présence d'une fraction B" obtenue au cours de la purification de TFIIB et contenant probablement des facteurs de transcription de l'ARN polymérase III impliqués dans l'initiation (Andrau & Werner, 2001). Il est à noter que des expériences de Western Blot ont révélé la présence de TFIIS dans cette fraction B". Nous avons pu observer que l'ajout de la protéine TFIIS sauvage restaure la production d'un transcrit unique de taille correcte, ce qui n'est pas le cas du mutant TFIIS-E291A. Des expériences d'extension d'amorce réalisées *in vitro* après un ou plusieurs cycles de transcription ont permis de confirmer que TFIIS agit sur le choix correct du site d'initiation de la transcription par l'ARN polymérase III. Ce résultat n'a cependant pas pu être reproduit *in vivo*, car probablement l'action d'autres protéines ayant un rôle similaire masque cet effet. Enfin, nous avons vérifié que TFIIS n'influence ni la cinétique d'élongation ni l'activité de clivage de l'ARN polymérase III.

L'ensemble de ces résultats indique un rôle inattendu de TFIIS dans la transcription par l'ARN polymérase III. Des données *in vivo* et *in vitro* suggèrent que TFIIS, plus particulièrement à travers son motif RSADE, est important dans le choix du site d'initiation de la transcription, et qu'il favoriserait le recrutement ou la stabilité de l'ARN polymérase III et de TFIIB au promoteur des gènes de classe III.

## Article





## Genome-wide location analysis reveals a role of TFIIIS in RNA polymerase III transcription

Yad Ghavi-Helm, Magali Michaut, Joël Acker, et al.

*Genes Dev.* 2008 22: 1934-1947

Access the most recent version at doi:[10.1101/gad.471908](https://doi.org/10.1101/gad.471908)

---

**Supplemental  
Material**

<http://genesdev.cshlp.org/content/suppl/2008/07/01/22.14.1934.DC1.html>

**References**

This article cites 52 articles, 26 of which can be accessed free at:  
<http://genesdev.cshlp.org/content/22/14/1934.full.html#ref-list-1>

**Email alerting  
service**

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#)

---

---

To subscribe to *Genes & Development* go to:  
<http://genesdev.cshlp.org/subscriptions>

---

# Genome-wide location analysis reveals a role of TFIIS in RNA polymerase III transcription

Yad Ghavi-Helm, Magali Michaut, Joël Acker, Jean-Christophe Aude, Pierre Thuriaux, Michel Werner,<sup>2</sup> and Julie Soutourina<sup>1</sup>

CEA, iBiTec-S, Gif-sur-Yvette Cedex F-91191, France

TFIIS is a transcription elongation factor that stimulates transcript cleavage activity of arrested RNA polymerase II (Pol II). Recent studies revealed that TFIIS has also a role in Pol II transcription initiation. To improve our understanding of TFIIS function *in vivo*, we performed genome-wide location analysis of this factor. Under normal growth conditions, TFIIS was detected on Pol II-transcribed genes, and TFIIS occupancy was well correlated with that of Pol II, indicating that TFIIS recruitment is not restricted to NTP-depleted cells. Unexpectedly, TFIIS was also detected on almost all Pol III-transcribed genes. TFIIS and Pol III occupancies correlated well genome-wide on this novel class of targets. *In vivo*, some *dst1* mutants were partly defective in tRNA synthesis and showed a reduced Pol III occupancy at the restrictive temperature. *In vitro* transcription assays suggested that TFIIS may affect Pol III start site selection. These data provide strong *in vivo* and *in vitro* evidence in favor of a role of TFIIS as a general Pol III transcription factor.

[**Keywords:** TFIIS; transcription; RNA polymerase III; RNA polymerase II; ChIP-chip]

Supplemental material is available at <http://www.genesdev.org>.

Received January 21, 2008; revised version accepted May 23, 2008.

Gene transcription is a complex and highly regulated process that, in eukaryotes, is carried out by three specialized RNA polymerases (Pol I, II, and III), dedicated to the transcription of different sets of genes. Transcription starts by the assembly of large preinitiation complexes (PIC) comprising TBP (the TATA-binding protein common to all three transcription systems) and general transcription factors specific to the RNA polymerase considered. Other transcription factors facilitate RNA elongation by Pol II (Shilatifard et al. 2003). One of these elongation factors, TFIIS, was initially discovered in human (Natori et al. 1973) and yeast cells (Sawadogo et al. 1980b), and is highly conserved in the eukaryotic and archaeal kingdoms (Hausner et al. 2000; Fish and Kane 2002). Structurally unrelated but functionally equivalent factors (GreA and GreB) also operate in bacteria (Borukhov et al. 1993; Opalka et al. 2003). TFIIS forms a binary complex with Pol II and stimulates an intrinsic transcript cleavage activity of RNA polymerase allowing elongating enzymes to resume RNA synthesis after accidental transcription arrest (Fish and Kane 2002; Kettenberger et al. 2003).

TFIIS is organized in three domains as determined by

limited proteolysis and nuclear magnetic resonance (NMR) (Awrey et al. 1997; Olmsted et al. 1998). The three-dimensional structure of the Pol II–TFIIS complex, without the TFIIS N-terminal domain, has been determined by crystallographic studies (Kettenberger et al. 2003). The N-terminal domain I of TFIIS protrudes from Pol II and is not required for elongation (Nakanishi et al. 1995; Awrey et al. 1998), but binds the yeast Mediator and SAGA coactivator (Wery et al. 2004). The TFIIS central domain II is inserted into the funnel-shaped pore of Pol II and brings the C-terminal conserved RSADE motif very close to the enzyme active site (Kettenberger et al. 2003). Remarkably, the C-terminal domain III of TFIIS shares the RSADE motif with the C-terminal parts of the Rpa12 (Pol I) and Rpc11 (Pol III) subunits, and is more distantly related to the corresponding Rpb9 subunit of Pol II, which has no RSADE motif. Rpc11 mediates the intrinsic transcription cleavage activity of Pol III (Chedin et al. 1998).

In the yeast *Saccharomyces cerevisiae*, TFIIS is encoded by *DST1*. The *dst1-Δ* mutant, or *dst1* mutants of the RSADE motif, grow like wild type, but are sensitive to the NTP-depleting effects of 6-azauracil (6AU) and mycophenolic acid (MPA), two drugs thought to compromise elongation efficiency (Exinger and Lacroute 1992; Ubukata et al. 2003). In the presence of 6-azauracil, deletion of *DST1* affects Pol II processivity without influencing its elongation rate (Mason and Struhl 2005). TFIIS was also

## Corresponding authors

<sup>1</sup>E-MAIL [julie.soutourina@cea.fr](mailto:julie.soutourina@cea.fr); FAX 33-1-69-08-47-12.

<sup>2</sup>E-MAIL [michel.werner@cea.fr](mailto:michel.werner@cea.fr); FAX 33-1-69-08-47-12.

Article is online at <http://www.genesdev.org/cgi/doi/10.1101/gad.471908>.

found to stimulate transcription past an artificial arrest site in vivo, thus further supporting the notion that it overcomes elongation blocks (Kulish and Struhl 2001). Recent reports suggest that, in addition to its well-documented role as a Pol II elongation factor, TFIIS also contributes to transcription initiation (Malagon et al. 2004; Wery et al. 2004; Prather et al. 2005; Guglielmi et al. 2007; Kim et al. 2007). This initiation role does not depend on the C-terminal RSADE motif needed for transcript cleavage stimulatory activity (Guglielmi et al. 2007; Kim et al. 2007).

Although the role of TFIIS in stimulating Pol II transcript cleavage activity is well characterized in vitro, its in vivo function remains less obvious. In this study, we analyzed the genome-wide location of TFIIS using chromatin immunoprecipitation coupled with DNA microarray hybridization (ChIP–chip). TFIIS was detected on many Pol II-transcribed genes under normal growth conditions, and its occupancy closely correlated with that of Pol II. Unexpectedly, we identified a novel class of TFIIS targets corresponding to Pol III-transcribed genes. In vitro Pol III transcription assays suggested that TFIIS may contribute to correct start site selection. Taken together, our in vivo and in vitro data reveal a previously unsuspected role of TFIIS in Pol III transcription.

## Results

### *TFIIS is located genome-wide on the Pol II- and Pol III-transcribed genes*

Previous models proposed that TFIIS protein was recruited during transcription elongation in conditions where Pol II was stalled. This hypothesis was strengthened by ChIP assays, suggesting that TFIIS does not normally reside on DNA but is specifically recruited in the presence of NTP-depleting drugs that favor the transcriptional arrest of Pol II elongation complexes (Pokholok et al. 2002). However, recent ChIP analyses revealed the presence of TFIIS on promoters and coding regions of several selected genes in cells grown under normal conditions (Prather et al. 2005; Guglielmi et al. 2007). To better understand TFIIS function and resolve this discrepancy, we performed genome-wide location analysis of TFIIS under standard conditions, or in cells that have been NTP-depleted in the presence of MPA.

We examined the genome-wide distribution of TFIIS and Pol II by ChIP–chip experiments and compared their relative enrichments. The DNA arrays used contained >40,000 oligonucleotide probes covering 12 Mb of the yeast genome (see the Materials and Methods). This analysis was done using a strain (YGH2) carrying an active N-terminal HA-tagged version of TFIIS. We found 3652 oligonucleotides significantly bound ( $P$ -value <0.005) by TFIIS under standard conditions. Correlation between TFIIS and Pol II enrichment showed an unexpected pattern (Fig. 1A). The distribution was split into two highly different data sets. A first set of oligonucleotides, located within 1419 Pol II-transcribed genes, showed a good correlation between TFIIS and Pol II enrichment. As dis-

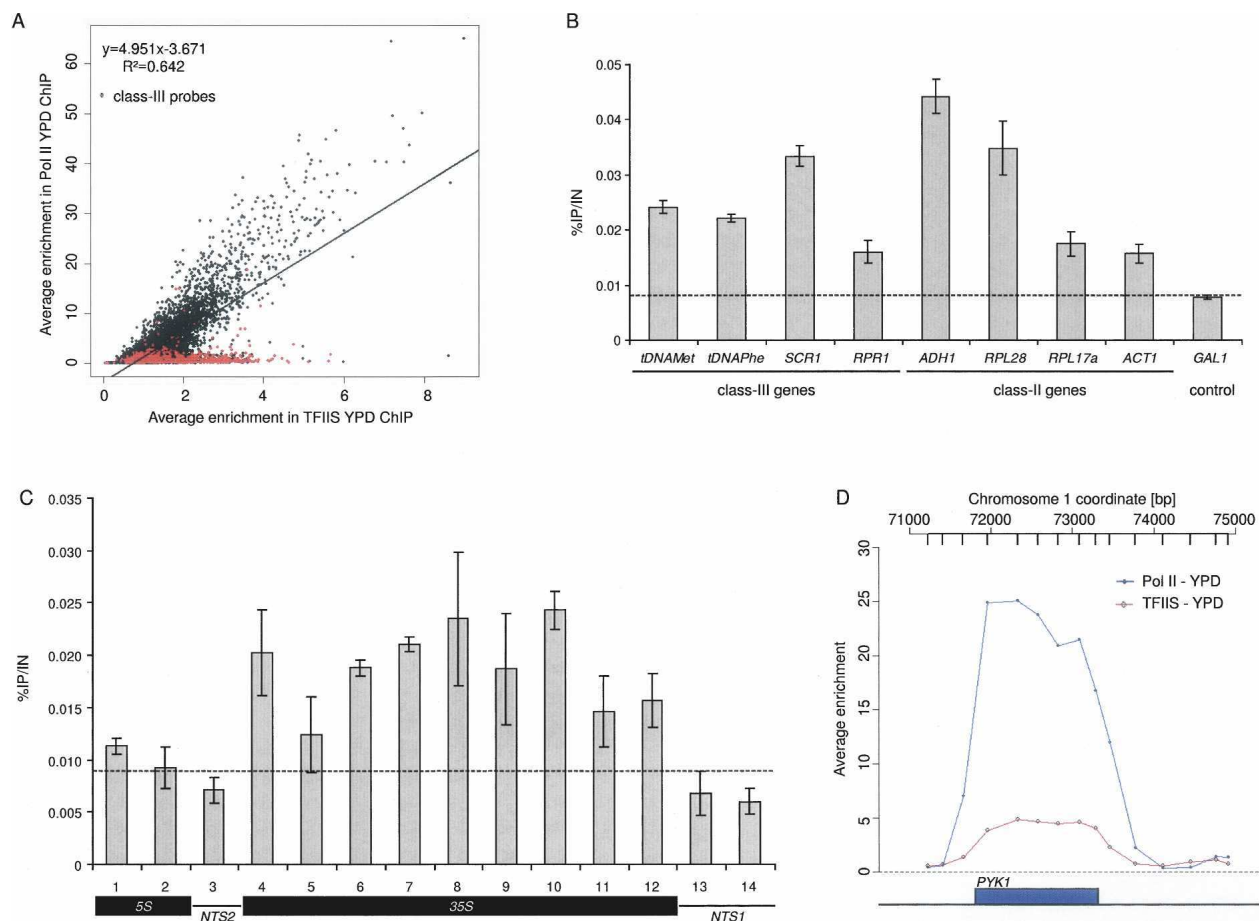
cussed below, a second group of oligonucleotides was enriched for TFIIS but not for Pol II, and corresponded to Pol III-transcribed genes.

The results obtained by ChIP–chip experiments were confirmed by conventional ChIP on a set of selected Pol II- and Pol III-transcribed genes (Fig. 1B). All selected genes displayed a significant enrichment compared to background level measured on the coding region of the nontranscribed *GAL1* gene. Similar results were obtained in ChIP experiments using chromatin from a strain expressing native TFIIS and anti-TFIIS antibodies for immunoprecipitation (data not shown). We wondered whether TFIIS could also be detected on Pol I-transcribed 35S rDNA and Pol III-transcribed 5S rDNA templates. Since the array used had a poor coverage of that region, we performed conventional ChIP experiments on the rDNA locus. A slight enrichment on the 35S rDNA and no significant enrichment on the 5S rDNA were seen compared to the intergenic NTS1 and NTS2 regions (Fig. 1C). Since the binding of TFIIS to transcribed sequences would also be observed if TFIIS was an RNA-binding protein, we tested if the enrichment of TFIIS was dependent on the presence of the transcribed RNA by digestion with RNase A/T1 in ChIP experiments. Digestion did not affect TFIIS enrichment (data not shown). We also examined TFIIS occupancy of a gene transcribed by a heterologous T7 RNA polymerase (Chen et al. 1987). We did not observe any significant TFIIS enrichment (Supplemental Fig. S1), indicating that TFIIS does not localize to transcribed sequences thanks to an RNA-binding property.

### *Genome-wide TFIIS occupancy on Pol II-transcribed genes correlates with Pol II occupancy*

Analysis of Pol II genome-wide occupancy revealed 15,911 oligonucleotides significantly bound ( $P$ -value <0.005). These probes corresponded to 3819 Pol II-transcribed genes. After removing all the oligonucleotides corresponding to Pol III-transcribed genes from the analysis, there was a fairly linear relationship between TFIIS and Pol II enrichment with a correlation coefficient of 0.642 (Fig. 1A). The example of Pol II and TFIIS enrichment profiles on the *PYK1* gene (Fig. 1D) confirms that TFIIS is present on coding regions even under standard growth conditions (i.e., in the absence of NTP-depleting drugs), and that TFIIS and Pol II occupancies are extremely well correlated. Other examples of enrichment profiles on *ADH1* and *PGK1* genes are shown in Supplemental Figure S2A,B. The presence of TFIIS is not limited to transcribed protein-coding regions, since we detected TFIIS on intergenic regions transcribed in short unstable RNAs. The examples of the cryptic transcript *NELO25C*, located on chromosome V between *RMD6* and *DLD3* genes (Wyers et al. 2005), and of *SRG1*, a transcriptional repressor of *SER3* (Martens et al. 2004), are shown in Supplemental Figure S2C,D. Pol II- and TFIIS-enriched genes were classified according to the biological process categories of gene ontology (GO). A standard hypergeometric test was used to determine the overrepresented





**Figure 1.** Genome-wide location analysis of TFIIS and Pol II. (A) Enrichment of Pol II versus enrichment of TFIIS. The genome-wide enrichment of Pol II and TFIIS in YPD medium was assessed from ChIP-chip experiments using YGH2 (3HA-TFIIS) strain. A linear regression for Pol II enrichment versus TFIIS enrichment on class II probes and its equation are indicated. Red dots correspond to class III probes. (B) Quantitative ChIP analysis of TFIIS enrichment on selected genes. Immunoprecipitated fragments from ChIP experiments were amplified with primers as indicated in the Supplemental Material. The *GAL1* ORF was used as a control. Error bars represent the standard deviation between at least three biological replicates. The background level was represented by a dotted line. (C) Quantitative ChIP analysis of TFIIS enrichment on rDNA. Immunoprecipitated fragments from ChIP experiments were amplified with primers as indicated in the Supplemental Material. The intergenic NTS1 and NTS2 regions were used as a control. Error bars represent the standard deviation between at least three biological replicates. The background level was represented by a dotted line. (D) Enrichment profile of Pol II and TFIIS on the *PYK1* gene. The enrichment of Pol II (blue) and TFIIS (red) on *PYK1* in YPD medium was assessed from ChIP-chip experiments using YGH2 (3HA-TFIIS) strain. The genomic positions of probe regions on chromosome 1 are indicated along the X-axis and represented by black points or circles. Watson strand-transcribed gene *PYK1* is colored in blue. The enrichment ratio is indicated along the Y-axis.

categories and associated *P*-values. Since the lists of Pol II and TFIIS overrepresented GO categories were largely overlapping (Supplemental Table S1), we concluded that TFIIS was not preferentially bound to a particular group of Pol II-transcribed genes. The smaller number of TFIIS-enriched genes compared to Pol II is best explained by the lower binding levels of TFIIS.

To explore the effect of NTP depletion on TFIIS genome-wide occupancy, yeast cells grown in minimal SD medium were exposed to 10  $\mu$ M MPA for 4 h (i.e., just when an effect on growth rate can be observed). We could detect a slight increase of TFIIS enrichment level compared to standard conditions (SD medium), but this effect was far less pronounced than in a previous study

using a different and nonfunctional tagged version of TFIIS (Supplemental Fig. S3A; Pokholok et al. 2002). The enrichment profiles of TFIIS on *PYK1* and *ILV5* genes illustrate that in the presence of mycophenolate, occupancy levels of TFIIS were only slightly increased all along the genes (Supplemental Fig. S3B,C). Note that TFIIS enrichment is much higher in yeast cells growing rapidly in rich medium as compared with minimal medium.

#### *Genome-wide TFIIS occupancy on Pol III-transcribed genes correlates with Pol III occupancy*

We compared TFIIS and Pol III occupancies on Pol III-transcribed genes, since a global correlation would imply

that TFIIS might be a new Pol III general factor. A ChIP-chip experiment on Rpc160, the largest Pol III subunit, was performed. We found 964 enriched oligonucleotides located in or close to the class III genes that were described previously in genome-wide location analyses (Harismendy et al. 2003; Roberts et al. 2003; Moqtaderi and Struhl 2004). Four-hundred-thirty-two of the 3652 oligonucleotides significantly bound ( $P$ -value  $<0.005$ ) by TFIIS under standard conditions were within or close (up to 500 bp upstream or downstream) to class III genes. They were significantly enriched by Pol III, and represented all Pol III-transcribed genes for which probes were located on the arrays; i.e., *tDNAs*, *SCR1*, *RPR1*, *SNR52*, and *ZOD1*. *SNR6* showed a low level of TFIIS enrichment on the B box located ~100 bp downstream from transcribed region (Supplemental Fig. S4A). No TFIIS enrichment was detected on the *ETC* loci that are occupied by TFIIB but not by TFIIC and Pol III (Moqtaderi and Struhl 2004). As noted above, the 5S rDNA also showed no detectable enrichment for TFIIS.

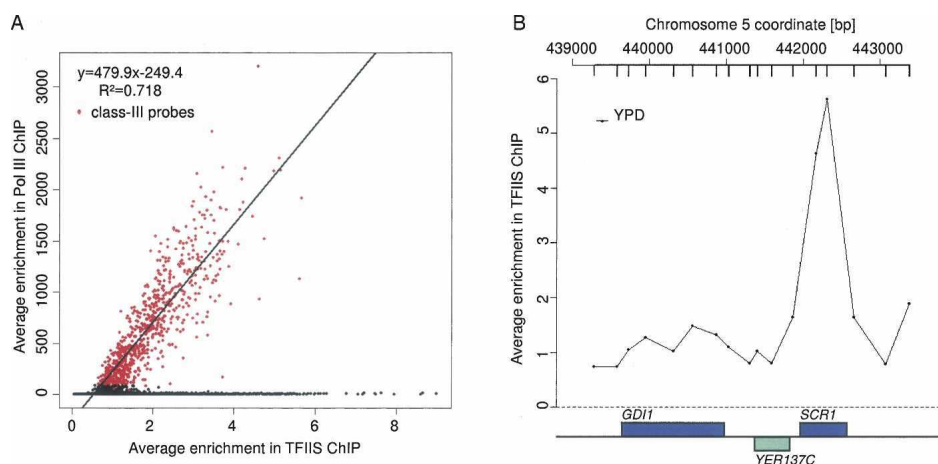
The overall correlation between Pol III and TFIIS occupancies (Fig. 2A) confirmed our previous observations on the genome-wide TFIIS distribution. Considering only the oligonucleotides corresponding to Pol III-transcribed genes, the analysis showed a linear relationship with a correlation coefficient of 0.718 (Fig. 2A). The enrichment profile of TFIIS on *SCR1*, a selected class III gene, is presented in Figure 2B. Other examples of TFIIS enrichment profiles on *RPR1* and *tDNA<sup>Gly</sup>* are shown in Supplemental Figure S4B,C. In these examples, the Pol II-transcribed genes adjacent to class III genes were devoid of TFIIS, strongly suggesting that TFIIS could be recruited independently of Pol II transcription.

To establish that the presence of TFIIS on class III genes was specific to Pol III transcription and independent of Pol II transcription, we used specific Pol II and

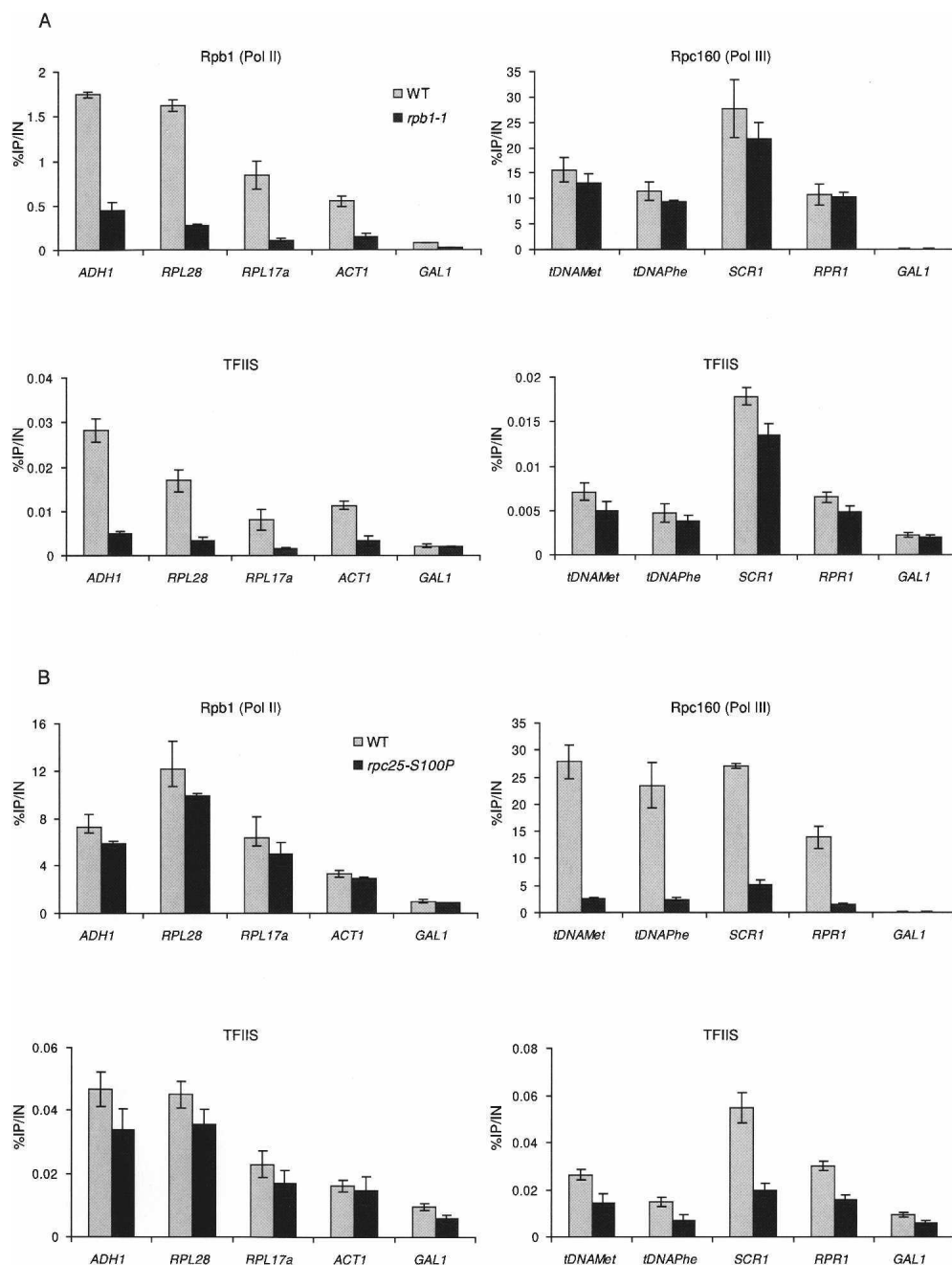
Pol III mutations and examined how they affected the distribution of TFIIS. The *rpb1-1* mutant of Pol II rapidly stops transcription after a shift to 37°C (Nonet et al. 1987). When cells were incubated at 37°C for 30 min, Pol II occupancy of all Pol II-transcribed genes tested (*ADH1*, *RPL28*, *RPL17a*, and *ACT1*) was greatly reduced in the *rpb1-1* strain compared to the wild-type strain (Fig. 3A, left panel). TFIIS occupancy strongly correlated with that of Pol II, and was also greatly reduced after incubation at 37°C. Conversely, TFIIS and Pol III occupancies of all Pol III-transcribed genes tested (*tDNAMet*, *tDNAPhe*, *SCR1*, and *RPR1*) remained largely unchanged under the same conditions (Fig. 3A, right panels). These results demonstrated that the presence of TFIIS on Pol III-transcribed genes is independent of Pol II transcription.

The Pol III-specific mutant *rpc25-S100P* was similarly used to impair class III gene transcription. For this purpose, a more prolonged shift to the restrictive temperature (10 h at 37°C) was needed (Zaros and Thuriaux 2005). Under this condition, Pol III and TFIIS occupancies were reduced on class III genes, even though the decrease of TFIIS occupancy was somewhat less pronounced (Fig. 3B, right panel). As expected, no significant effect was observed on Pol II and TFIIS association to the Pol II-transcribed genes (Fig. 3B, left panel). Thus, the presence of TFIIS on class II or III genes is largely dependent on transcription by Pol II or Pol III, respectively.

Most class III genes are too small to allow a spatial resolution of the Pol III machinery location by ChIP, but *SCR1*, the longest Pol III-transcribed gene (522 bp), provided this possibility. We examined the spatial distribution of TFIIS on this gene compared with that of the Pol III basal machinery composed, in addition to Pol III, of TFIIB and TFIIC complexes (Geiduschek and Kassavetis 2001). Seven real-time PCR amplicons spanning the upstream, transcribed, and downstream regions of



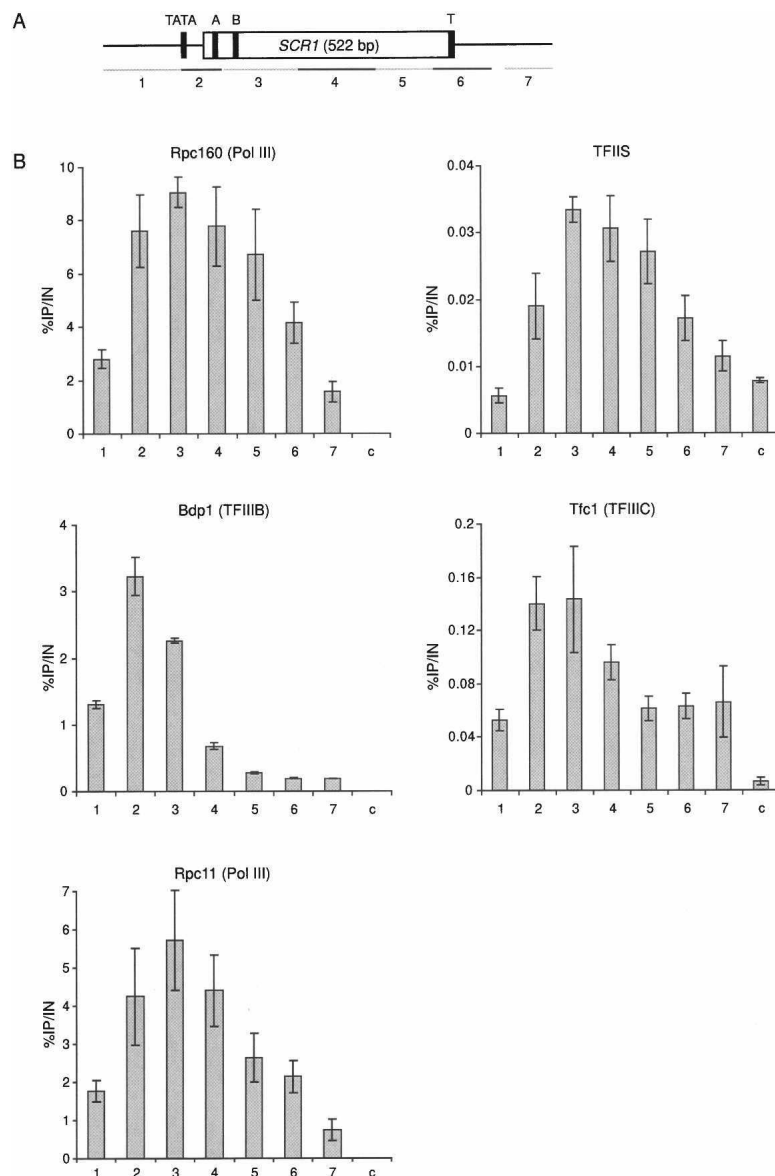
**Figure 2.** Genome-wide location analysis of TFIIS and Pol III. (A) Enrichment of Pol III versus enrichment of TFIIS. The genome-wide enrichment of Pol III and TFIIS in YPD medium was assessed from ChIP-chip experiments with MW671 (3HA-RPC160) and YGH2 (3HA-TFIIS) strains, respectively. A linear regression and its equation are indicated. Red dots correspond to class III probes. (B) Enrichment profile of TFIIS on the *SCR1* gene. The genome-wide enrichment of TFIIS in YPD medium was assessed from ChIP-chip experiments. The genomic positions of probe regions on chromosome 5 are reported along the X-axis and represented by black points. Watson strand-transcribed genes are colored in blue and Crick strand-transcribed genes are colored in green. The enrichment ratio is reported along the Y-axis.



**Figure 3.** Effect of *rpb1-1* and *rpc25-S100P* mutations on TFIIS, Pol II, and Pol III occupancies on selected genes. Immunoprecipitations were performed using antibodies against 3HA (12CA5) for HA-TFIIS, 13Myc (9E10) for Rpc160-13Myc, and CTD (8WG16) for the Pol II Rpb1 subunit. Immunoprecipitated fragments from ChIP experiments were amplified with primers as indicated in the Supplemental Material. The *GAL1* ORF was used as a control. The values are the average of three independent experiments. Error bars indicate the standard deviation. (A) TFIIS, Pol II, and Pol III occupancies in the *rpb1-1* mutant. Standard ChIP assays were performed on chromatin prepared from the D788-4a strain transformed by *RPB1*-containing pYeB-B220 plasmid (wild type) or by empty vector Yep351 (*rpb1-1*). Cells were grown in selective SD medium complemented with amino acids at 30°C and then shifted for 30 min at 37°C. (B) TFIIS, Pol II, and Pol III occupancies in the *rpc25-S100P* mutant. Standard ChIP assays were performed on chromatin prepared from D792-3a strain transformed by *RPC25*-containing pRS315-RPC25 plasmid (wild type) or by empty vector pRS315 (*rpc25-S100P*). Cells were grown in selective SD medium complemented with amino acids at 30°C and then shifted for 10 h at 37°C.

*SCR1* were designed (Fig. 4A). We analyzed the association profiles of the Rpc160 Pol III subunit, the Bdp1 TFIIB subunit, the Tfc1 TFIIC subunit, and TFIIS (Fig.

4B). The distribution of the Pol III transcription machinery on *SCR1* was as described previously (Roberts et al. 2006). TFIIB binding was maximal on the TATA box,



**Figure 4.** Pol III, TFIIS, TFIIB, and TFIIC location analysis on the *SCR1* gene. (A) Schematic organization of the *SCR1* gene. The location of the PCR fragments amplified in ChIP analyses are indicated by gray and black lines. Black boxes represent TATA, A and B boxes, and Terminator (T). (B) Occupancy profile of Pol III, TFIIS, TFIIB, and TFIIC on the *SCR1* gene. Standard ChIP assays were performed on chromatin prepared from YGH11 strain using antibodies against 3HA-TFIIS (12CA5) and Rpc160-13Myc (9E10), from MW4035 strain using antibodies against Bdp1-3HA (12CA5), from yOH1 strain using antibodies against Tfc1-13Myc (9E10), and from YGH15 strain using antibodies against Rpc11-3HA (12CA5). Cells were grown in YPD medium at 30°C. The *GAL1* ORF was used as a control [c].

while TFIIC was cross-linked over the entire *SCR1* locus. TFIIS distribution on the *SCR1* gene resembled more closely that of Pol III with a maximum at the level of the A and B boxes at the beginning of the gene. However, the TFIIS occupancy profile differed from those of TFIIB and TFIIC. We also examined the association of Rpc11, a Pol III subunit homologous to TFIIS, and showed that the distribution of Rpc11 on the *SCR1* gene was similar to that of Rpc160 subunit (Fig. 4B). Thus, the TFIIS distribution on the *SCR1* gene follows closely that of Pol III, suggesting an active role of TFIIS in class III-gene transcription.

Since TFIIS interacts directly with Pol II, we tested the possible association of this factor with Pol III machinery by coimmunoprecipitation (co-IP) approach. No interaction of TFIIS with Pol III, TFIIC, and TFIIB was detected in crude extracts from noncross-linked cells (Supplemental Fig. S5A). However, when co-IP assays were performed with extracts from cross-linked cells (ChIP extracts), TFIIS

coimmunoprecipitated with Pol III and TFIIC, but not with TFIIB (Supplemental Fig. S5B), in agreement with our ChIP data. We detected TFIIS when Rpc160-Myc or Rpc11-HA were used to immunoprecipitate Pol III. These co-IP results showed a concomitant presence of TFIIS with Pol III and TFIIC on class III genes.

#### Effect of *dst1* mutations on tRNA synthesis and Pol III occupancy in vivo

Since tRNA suppressor genes are transcribed by Pol III, we tested the influence of deleting *DST1* on suppression efficiency to analyze the role of TFIIS in Pol III transcription in vivo. We therefore compared the growth of a yeast strain containing the chromosomally integrated *SUP11* tRNA<sup>Tyr</sup><sub>UAA</sub> allele suppressing the *ade2-1* ochre mutation with that of the isogenic *dst1-Δ* strain. In the presence of adenine in the growth medium, the *SUP11*



*dst1-Δ ade2-1* strain grew as well as the *SUP11 DST1 ade2-1* strain (Fig. 5A, +Ade), but a significant growth difference was observed between the two strains in the medium lacking adenine (–Ade), indicating a decrease in the level of *ade2-1* mutation suppression. This observation suggests that TFIIS could influence Pol III transcription *in vivo*.

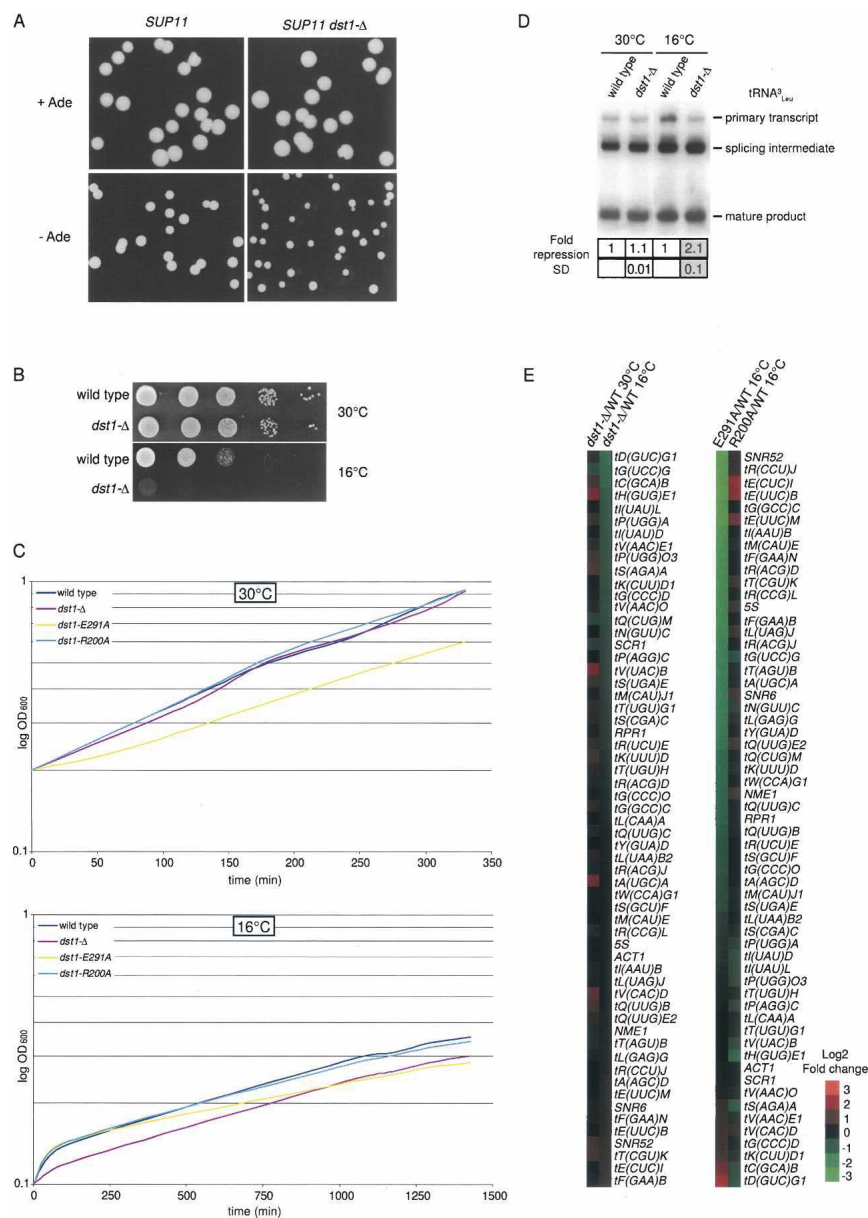
In the course of this study, we observed that *dst1-Δ* impaired yeast growth at 16°C (Fig. 5B), a phenotype that has not been described previously. This phenotype was not complemented by a point mutation in the RSADE motif of TFIIS domain III, responsible for the cleavage activity (*dst1-E291A*) (Fig. 5C). This mutant, the most sensitive to cold temperature, actually had a growth defect even at 30°C. In contrast, the *dst1-R200A* mutant in Pol II-binding domain II grew like a wild type. We con-

cluded that the domain III of TFIIS is important for yeast growth at low temperature.

We analyzed the effect of *dst1* deletion on the levels of short-lived tRNA<sub>3</sub><sup>Leu</sup> precursor at low temperature. Wild-type strain and a *dst1-Δ* mutant were grown at 30°C or shifted for 8 h at 16°C. Total RNAs were extracted and used for Northern blot analysis (Fig. 5D). The amount of mature tRNA<sub>3</sub><sup>Leu</sup> was used as a loading control. The deletion of *dst1* significantly decreased the tRNA<sub>3</sub><sup>Leu</sup> precursor levels in yeast cells grown at 16°C (2.1-fold).

To investigate the effect of *dst1* mutations on class III gene expression under low-temperature conditions (16°C) in more details, we performed a genome-wide analysis of Pol III transcriptome (Ciesla et al. 2007). Log<sub>2</sub> ratios of class III gene transcription levels in the mutant versus

**Figure 5.** In vivo effect of TFIIS on Pol III transcription. (A) Growth of the YGH10 (*SUP11 dst1-Δ ade2-1 ochre*) strain relative to the MB159-2c (*SUP11 DST1 ade2-1 ochre*) strain. Cells were grown at 30°C for 3 d on minimal medium supplemented (+Ade) or not (–Ade) with adenine. (B) Growth of the *dst1*-null mutant at 30°C or 16°C. Suspensions of YPH500 (wild-type) or ESH1 (*dst1-Δ*) strains were serially diluted, spotted on YPD-rich medium, and grown for 3 d at 30°C or at for 5 d 16°C. (C) Growth of *dst1* mutants at 30°C or 16°C. *dst1-Δ* (YGH12) strain was transformed by pRS425-pPGK vector, pRS425-DST1, pRS425-*dst1R200A*, or pRS425-*dst1E291A* plasmids bearing different *DST1* alleles. Cells were grown in liquid SC-leucine medium at 30°C or shifted at time 0 at 16°C. (D) Northern blot analysis of tRNA<sub>3</sub><sup>Leu</sup> abundance in wild-type and *dst1-Δ* strains. RNAs were extracted from wild-type (ESH1/pRS425-DST1) and *dst1-Δ* (ESH1/pRS425-pPGK) cells grown in SC-leucine medium at 30°C or shifted at 16°C for 8 h. tRNA precursors abundance was quantified, and fold decrease of transcription in *dst1-Δ* strain was compared with that in wild type grown at the same temperature (arbitrary set to 1). Average values of three independent experiments and standard deviations (SD) are indicated. (E) Analysis of the Pol III transcriptome at 16°C. RNAs were extracted from the *dst1-Δ* (ESH1) strain transformed by pRS425-pPGK vector, pRS425-DST1, pRS425-*dst1R200A*, or pRS425-*dst1E291A* plasmids grown at 30°C or shifted for 8 h to 16°C, and analyzed by hybridization to Pol III-specific microarrays. Expression log<sub>2</sub> ratios of class III genes in the mutant strains compared with wild type are presented according to the red-green color scale.



wild-type strains were hierarchically clustered (Fig. 5E). At 30°C, the Pol III transcript profiles of the *dst1-Δ* and the wild-type strains were similar. In contrast, a significant decrease of several class III transcripts was observed in *dst1-Δ* mutant at 16°C. We found that transcription levels of class III genes were strongly diminished at 16°C in the *dst1-E291A* strain, whereas the *dst1-R200A* mutant had a Pol III transcript profile similar to that of the wild type. The *dst1-E291A* mutant had more pronounced effects on Pol III transcription than the *dst1-Δ* strain, consistent with its slower growth at all temperatures. Curiously, class III genes showing the most reduced transcript levels were different in *dst1-Δ* and *dst1-E291A* mutants.

To examine the effect of the *dst1-Δ* mutation on genome-wide Pol III occupancy at low temperature (16°C), wild-type and *dst1-Δ* strains containing a C-terminal 13Myc tag on the Rpc160 Pol III subunit were grown in YPD-rich medium at 30°C and then shifted to 16°C for 8 h. At low temperature, all class III genes were bound by Pol III. We compared Pol III genome-wide occupancy in a *dst1-Δ* and a wild-type strain grown at 16°C (Fig. 6A). A general reduction of Pol III occupancy in the *dst1-Δ* mutant was observed, suggesting that TFIIS could stabilize Pol III on class III genes. Regression analysis indicated a 1.5-fold decrease of Pol III binding in the mutant (the slope of the correlation line, shown in red, equal to 0.645) with a high correlation coefficient ( $R^2 = 0.965$ ). The high correlation coefficient indicated that the association of Pol III to almost all class III genes was significantly reduced. A similar analysis was performed to examine the effect of *dst1* deletion on Pol II genome-wide occupancy at low temperature (Fig. 6B). The binding of Pol II was found to be reduced 1.6-fold in the mutant (the slope of the correlation line equal to 0.621) with a reduced correlation between wild-type and *dst1-Δ* strains ( $R^2 = 0.647$ ). In the case of Pol II, the *dst1* deletion could have indirect effects on gene expression regulation that may explain the lower correlation coefficient between the wild-type and the mutant strains. Examples of enrichment profiles on specific genes are shown in Supplemental Figure S6. The overall reduction of Pol II and Pol III occupancies that was observed at low temperature when *dst1* was deleted suggested that TFIIS could stabilize Pol II and Pol III on their target genes.

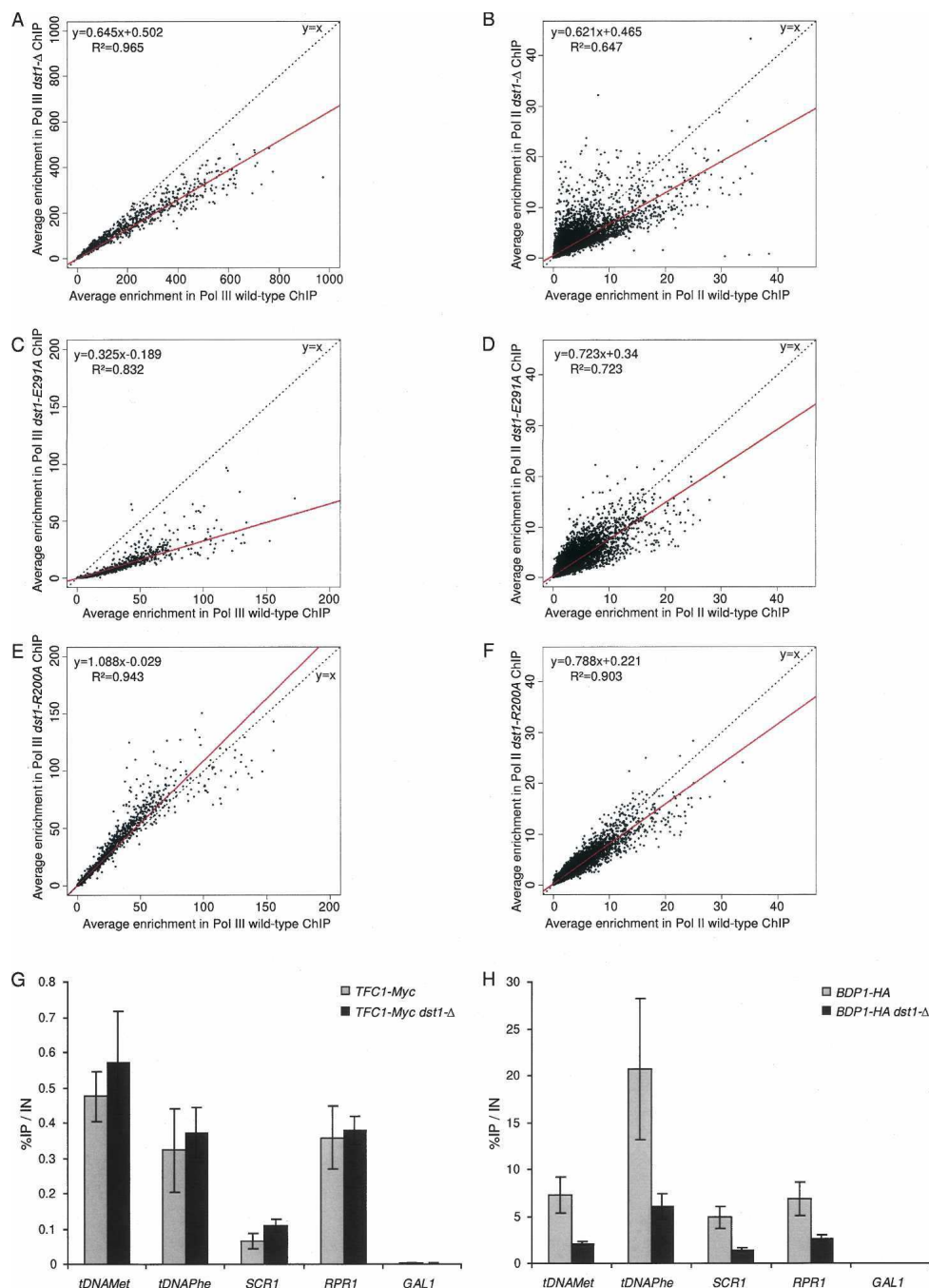
We extended our genome-wide Pol III and Pol II location analysis to the *dst1-E291A* and *dst1-R200A* point mutants grown at 16°C. The binding of Pol II was found to be reduced 1.4-fold in the *dst1-E291A* mutant and 1.3-fold in the *dst1-R200A* mutant (the slope of the correlation lines equal to 0.723 and 0.788, respectively) (Fig. 6D,F). As expected, both *dst1* mutations affected Pol II occupancy. In contrast, the binding of Pol III was significantly diminished (3.1-fold) in the *dst1-E291A* mutant, but not at all in the *dst1-R200A* mutant (the slope of the correlation lines equal to 0.325 and 1.088, respectively) (Fig. 6C,E). Thus, the *dst1-E291A* mutant affected both Pol II and Pol III binding, whereas the *dst1-R200A* mutant had a small but significant effect only on Pol II association.

We further analyzed the occupancy of the Bdp1 TFIIB subunit and the Tfc1 TFIIC subunit in *dst1-Δ* and wild-type strains on several class III genes at 16°C (Fig. 6G,H). The association of TFIIC was unchanged, but the binding of TFIIB was reduced in the *dst1-Δ* mutant compared with the wild type.

#### TFIIS affects Pol III transcription in vitro

To examine the role of TFIIS in Pol III transcription in vitro, we expressed wild-type TFIIS and the TFIIS-E291A mutant form that is unable to stimulate Pol II elongation (Ubukata et al. 2003) as 6xHis fusion proteins in *Escherichia coli*, and purified the corresponding polypeptides to near homogeneity. We first checked that the wild-type TFIIS was active in stimulating nonspecific Pol II transcription on calf thymus DNA (Sawadogo et al. 1980b), while the TFIIS-E291A protein was not (data not shown). The effect of the wild-type and the E291A mutant TFIIS on Pol III transcription in vitro were examined in a Pol III transcription system reconstituted with all recombinant TFIIC and TFIIB and highly purified Pol III (Ducrot et al. 2006). Multiple-round transcription assays were performed with the *SUP4* tRNA gene as a template. As observed previously, the *SUP4* transcripts generated with TFIIB recombinant components migrated as two or three diffuse bands on polyacrylamide gels (Fig. 7A, lane 1), which was not the case in the presence of the crude fraction B' (Andrau and Werner 2001). This fraction contains additional factors, like Nhp6 (Braglia et al. 2007), that can restore transcriptional initiation specificity (Kassavetis and Steiner 2006). Western blotting analysis also revealed the presence of TFIIS in the B' fraction (data not shown). Remarkably, adding the purified wild-type TFIIS to the reconstituted transcription system resulted in the formation of the correct length transcript (Fig. 7A, lanes 2–6), in contrast to the purified TFIIS-E291A mutant protein at the same concentration (Fig. 7A, lanes 7–11). The same results were obtained after mixing wild-type and mutant TFIIS preparations, indicating that the mutant TFIIS preparation did not contain interfering components (Fig. 7A, lane 12). Quantification of the transcription signals revealed that the total amount of transcripts did not significantly change in the presence of wild-type or E291A mutant TFIIS. High concentrations of TFIIS (40 ng/μL) started to inhibit the transcription (Fig. 7A, lanes 6–11). In the subsequent assay, we used 5 ng/μL TFIIS.

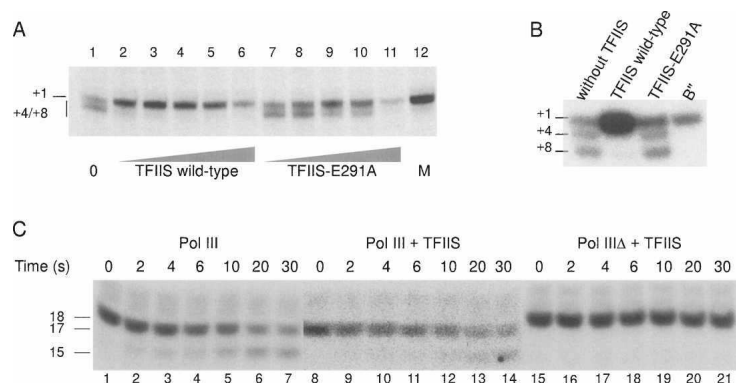
To demonstrate that TFIIS could contribute to correct start site selection, RNAs generated in a standard multiple-round Pol III transcription assay were purified and analyzed by primer extension. Figure 7B showed that, in the absence of TFIIS, transcription initiation occurred at base pairs +1, +4, and +8 relative to the start site used in vivo. The addition of B' fraction or of wild-type TFIIS restored a correct start site selection, while the TFIIS-E291A mutant had no effect. Similar results were obtained on transcripts produced in a single-round assay (Supplemental Fig. S7A). The transcription start sites were analyzed in vivo on *SUP4* gene in a *nhp6a-Δ nhp6b-Δ* con-



**Figure 6.** Pol II and Pol III genome-wide enrichment in TFII mutants at 16°C. The genome-wide enrichment of Pol III (A,C,E) or Pol II (B,D,F) in YPD medium (A,B) or SC-leucine medium (C–F) at 16°C was assessed from ChIP–chip experiments. A linear regression (red line) and its equation are indicated. The dotted line corresponds to  $y = x$ . Enrichment of Pol III or Pol II in *dst1-Δ* (YGH12) was compared with the wild-type strain (YGH11) (A,B). Enrichment of Pol III or Pol II in *dst1-E291A* (YGH12/pRS425-*dst1E291A*) (C,D) and *dst1-R200A* (YGH12/pRS425-*dst1R200A*) (E,F) strains was compared with the wild-type strain (YGH12/pRS425-DST1). (G,H) Quantitative ChIP analysis of TFIIC and TFIIB enrichment on selected class III genes. *dst1-Δ* and wild-type strains were grown in YPD medium at 30°C and shifted for 8 h to 16°C. Immunoprecipitations were performed using antibodies against 13Myc (9E10) for Tfc1-13Myc and 3HA (12CA5) for Bdp1-3HA. Immunoprecipitated fragments from ChIP experiments were amplified with primers as indicated in the Supplemental Material. The *GAL1* ORF was used as a control. Error bars represent the standard deviation between at least three biological replicates.

text and in the presence or absence of *DST1* gene. The *nhp6* mutant background was used because these proteins were previously implicated in start site selection and

could mask the effect of *dst1-Δ*. We could not identify any effect of *dst1* deletion on start site selection in this background (Supplemental Fig. S7B).



**Figure 7.** TFIIS stimulates faithful Pol III transcription in vitro. (A) Effect of TFIIS on Pol III transcription. Standard in vitro Pol III transcription on the *SUP4* template has been performed as described in the Materials and Methods, in the absence (lane 1) or presence of increasing quantities (0.1, 0.2, 0.4, 0.8, and 1.6  $\mu$ g) of wild-type TFIIS (lanes 2–6) or mutant TFIIS–E291A (lanes 7–11) or with a mix of wild-type and E291A protein (M, lane 12). Transcription start sites are indicated. (B) Primer extension analysis of start sites used by Pol III on the *SUP4* gene in vitro. Transcription reactions were performed in the absence of TFIIS, in the presence of wild-type TFIIS, mutant TFIIS–E291A, or with a B' fraction. Transcription and primer extensions were performed as described in Materials and Methods using a probe hybrid-

izing within the *SUP4* intron. Positions of the major transcription start site are indicated. (C) Time-course analysis of RNA cleavage by Pol III. Pol III (lanes 1–14) or Pol III $\Delta$  (lanes 15–21) (Chedin et al. 1998) ternary complexes formed in the presence of 3XTPs were isolated on Sepharose CL-2B as described in the Supplemental Material and then incubated for 10 min with a 50 M excess of purified TFIIS (lanes 8–21) or without TFIIS (lanes 1–7). Ternary complexes were then incubated at 16°C in transcription buffer containing 5 mM MgCl<sub>2</sub> in the absence of nucleotides for various periods of time. The transcript sizes are indicated. Transcription with Pol III $\Delta$  results in the formation of an 18-mer RNA instead of a 17-mer for the wild-type enzyme (Chedin et al. 1998).

To examine whether TFIIS could affect the elongation step of Pol III transcription, the elongation kinetics of Pol III were analyzed on a *SUP4* template that can produce a stalled ternary complex after the synthesis of a 17-mer transcript in the absence of GTP. The ternary complex was allowed to resume transcription by adding the four NTPs. No significant changes in elongation kinetics were observed in the presence or absence of purified wild-type TFIIS (Supplemental Fig. S8). TFIIS being a cleavage stimulatory factor in Pol II transcription, we wanted to know whether TFIIS could influence the intrinsic cleavage activity of Pol III. Labeled ternary complexes halted by omission of GTP in the transcription reaction were purified and incubated for various periods of time in the presence of MgCl<sub>2</sub> to activate the Pol III cleavage activity (Chedin et al. 1998). TFIIS did not stimulate the cleavage activity of wild-type Pol III (Fig. 7C, lanes 1–14, look at the disappearance rate of the 17-mer transcript). Pol III $\Delta$ , a RNA polymerase mutant that lacks the Rpc11, Rpc37, and Rpc53 subunits, is not competent for RNA cleavage activity (Chedin et al. 1998; Landrieux et al. 2006). Pol III $\Delta$  was purified from *rpc37HA-Ct* mutant as described previously (Landrieux et al. 2006). We observed that the addition of TFIIS did not restore an efficient cleavage by Pol III $\Delta$  (Fig. 7C, lanes 15–21). Taken at face value, in vitro transcription assays suggest a role for TFIIS in start site selection during Pol III transcription.

## Discussion

In this study, we performed a genome-wide location analysis of the TFIIS transcription factor. TFIIS was detected across the whole genome of exponentially growing cells, indicating that the binding of this transcription factor to chromatin is not restricted to NTP-depleted cells. TFIIS and Pol II genome-wide occupancies correlated very well, suggesting that TFIIS is not recruited only when Pol II is stalled. A second and more surprising outcome of our study was that TFIIS could be detected

on almost all Pol III-transcribed genes. This result raised the intriguing possibility that TFIIS might operate as a general Pol III-associated factor. We provided substantial in vivo and in vitro data demonstrating that TFIIS is important for Pol III transcription.

Concerning the TFIIS function in Pol II transcription, we observed that TFIIS was associated with a large number of Pol II-transcribed genes in cells growing exponentially in rich medium, in line with previous ChIP analyses of TFIIS on a few class II genes (Prather et al. 2005; Guglielmi et al. 2007). The distribution of TFIIS over distinct Pol II-transcribed regions, including intergenic regions transcribed in short unstable RNAs (Martens et al. 2004; Wyers et al. 2005; Davis and Ares 2006), precisely correlated with Pol II itself, which is consistent with the role of TFIIS as an elongation factor but is not contradictory with an additional role during initiation (Prather et al. 2005; Guglielmi et al. 2007; Kim et al. 2007). In the presence of MPA, the enrichment level of TFIIS genome-wide was only slightly increased compared with that in normal growth condition. Essentially, two models for TFIIS recruitment to Pol II-transcribed genes may be envisioned. TFIIS could be only recruited to arrested Pol II complexes. Alternatively, it might associate and dissociate from the elongating Pol II, independently of transcription arrests, shifting to the cleavage-prone conformation of Pol II if need arises. The latter model would better account for the fact that GTP depletion by MPA, which is likely to promote arrest, does not strongly increase TFIIS occupancy.

The presence of TFIIS at nearly all Pol III-transcribed genes strongly suggests that it is a Pol III transcription factor. This hypothesis is supported by several lines of evidence. (1) TFIIS occupancy correlated well with that of Pol III genome-wide. (2) The occupancy profile of TFIIS closely followed that of Pol III on the *SCR1* gene, the longest class III gene. (3) TFIIS coimmunoprecipitated with Pol III after formaldehyde cross-linking. (4) A temperature-sensitive Pol II mutation (*tpb1-1*) strongly



reduced TFIIS enrichment at Pol II-transcribed genes but had no effect on its association with Pol III-transcribed genes. Conversely, a temperature-sensitive Pol III mutation (*rpc25-S100P*) diminished TFIIS binding at Pol III-transcribed genes with no effect on its association with Pol II-transcribed genes. Thus, the presence of TFIIS on Pol III-transcribed genes depends on Pol III activity and is independent of Pol II. (5) Under low-temperature conditions, the *dst1-Δ* mutation affected growth, and diminished Pol III and TFIIB association with class III genes. (6) The *dst1* deletion impaired the translational suppression of *ade2-1* (a nonsense UAA mutant) by the *SUP4<sup>ochre</sup>* suppressor, and reduced the Pol III transcript levels under low-temperature conditions. (7) TFIIS improved Pol III transcription start site selection in vitro. Altogether, these data indicate that TFIIS is a bona fide component of the Pol III transcription machinery. It has been suggested previously that TFIIS could also play a role in Pol I transcription (Sawadogo et al. 1980a; Sawadogo et al. 1981; Schnapp et al. 1996), but a recent report demonstrated that the intrinsic cleavage activity of Pol I requires the Rpl2 subunit, sharing sequence homology with TFIIS (Kuhn et al. 2007). Our ChIP results showed that TFIIS was enriched on the Pol I-transcribed rDNA templates. TFIIS could thus be implicated in Pol I transcription. Interestingly, the Pol III-transcribed 5S rRNA genes, arranged in tandem with the 35S rRNA genes, were not enriched by TFIIS. The reason for the absence of TFIIS on 5S rRNA genes is unknown but could stem from the intragenic binding of TFIIB factor that is required for 5S transcription.

As a Pol II elongation factor, TFIIS strictly depends on the RSADE domain, since inactivating this domain (or deleting the entire TFIIS) makes cells sensitive to NTP-depleting drugs (mycophenolate, 6-azauracil), which is generally seen as a consequence of a defective Pol II-associated cleavage (Exinger and Lacroute 1992; Ubukata et al. 2003). As a factor involved in Pol II initiation, TFIIS is needed for the full recruitment of Pol II to several promoters, especially in the absence of the Med31 subunit of Mediator. This, however, does not depend on the RSADE motif but is impaired in *dst1-R200A*, a mutation of the TFIIS Pol II-interacting domain (Guglielmi et al. 2007). Turning now to the Pol III-associated role(s) of TFIIS, we found that the RSADE motif is clearly important in this context, since the corresponding mutant form of TFIIS alters start site selection by Pol III in vitro, and since Pol III occupancy is strongly diminished in *dst1-E291A* cells grown at 16°C, with a strongly perturbed Pol III transcriptome. The reason why class III genes, the transcription of which is most diminished, differ in *dst1-Δ* and *dst1-E291A* is presently unknown but might be the consequence of the strong perturbation of Pol III transcription in the latter background. An indirect effect of altered Pol I or Pol II transcription is unlikely because *dst1* mutations do not affect 35S rRNA precursor transcription at 16°C (data not shown) and because Pol II occupancy defect is not increased in *dst1-E291A* compared with *dst1-Δ*. In contrast, *dst1-R200A* had a limited but significant effect on Pol II occupancy, with no effect

at all on Pol III, suggesting that this mutation might specifically affect a Pol II-associated function (Guglielmi et al. 2007).

In the course of this study, we examined the possible implication of TFIIS in the different steps of in vitro Pol III transcription. Omitting TFIIS did not detectably influence Pol III elongation or cleavage activity in vitro but altered start site selection, suggesting that TFIIS might primarily act at the level of Pol III recruitment and/or transcription initiation. This is further supported by our ChIP assays showing that *dst1-Δ* strongly reduces the presence of Bdp1, the TFIIB-related subunit of the TFIIB initiation factor of Pol III, with no effect on the TFIIC initiation factor. In Pol III transcription, start site selection and initiation require the precise targeting of the enzyme by its initiation factors (TFIIC and TFIIB) and the opening of the transcription bubble around the start site (Kassavetis and Geiduschek 2006). TFIIS might participate in the initial steps of the Pol III transcription cycle in different ways. TFIIS could bend DNA and facilitate appropriate DNA binding by the transcription machinery as do Nhp6A and Nhp6B (Kassavetis and Steiner 2006). Alternatively, TFIIS could facilitate productive initiation by influencing the interaction of Pol III with basal factors for more accurate enzyme positioning. We favor the second hypothesis since direct binding of TFIIS to DNA has never been shown and since TFIIS and Pol III coimmunoprecipitated in cross-linked extracts.

One could be initially surprised to find TFIIS associated with class III genes, since Pol III has an intrinsic transcript cleavage activity that depends on Rpl1 subunit (Chedin et al. 1998; Alic et al. 2007). Rpl1 has a C-terminal Zn loop that bears an RSADE motif, critical for transcript cleavage, that closely resembles the C-terminal domain of TFIIS (Chedin et al. 1998). Our results suggest that both Rpl1 subunit and TFIIS are required for efficient Pol III transcription, but that their roles are not identical. Recombinant TFIIS (Fig. 7), in contrast to recombinant Rpl1 alone (Chedin et al. 1998), could not restore cleavage activity of Pol IIIΔ variant lacking Rpl1, suggesting that TFIIS does not participate in this reaction. Further, we found that Rpl1 and TFIIS bound throughout *SCR1* gene as Rpl160 (used as a proxy for Pol III) did. Moreover, TFIIS and Rpl1 coimmunoprecipitated in cross-linked extracts, as did Rpl160, indicating that their binding to class III genes is not mutually exclusive.

In conclusion, there is now mounting evidence that TFIIS controls several levels of DNA transcription. At a subset of Pol II-transcribed gene promoters, it could be recruited and act together with Mediator to recruit Pol II (Prather et al. 2005; Guglielmi et al. 2007; Kim et al. 2007), independently of its transcript cleavage stimulatory activity. TFIIS could also dynamically associate and dissociate from Pol II and stimulate the enzyme intrinsic RNA cleavage activity when needed. In addition, the present study shows that TFIIS is a Pol III transcription factor that stimulates Pol III transcription and may contribute to precise start site selection.

## Materials and methods

### Protein purification

DB3.1 *E. coli* cells containing either pDEST17-*DST1* or pDEST17-*dst1E291A* were grown at 30°C to 0.6 OD<sub>600</sub>. Expression of the 6xHis-TFIIS fusion protein was induced by addition of 1 mM isopropyl-1-thio-β-D-galactopyranoside. Cells were harvested after 3 h of induction and resuspended in 20 mM HEPES buffer (pH 7.5) containing 10 μM ZnCl<sub>2</sub>, 300 mM NaCl, 10% glycerol, 10 mM β-mercaptoethanol, and a set of protease inhibitors (phenyl-methyl-sulfonyl fluoride, Complete [Roche]). After lysis by sonication at 4°C, the lysate was clarified by centrifugation in a Beckman JA20 rotor for 25 min at 12,000 rpm. The 6xHis-TFIIS proteins were purified using an ÄKTA purifier (Amersham Biosciences) on a Hi Trap Chelating HP 5-mL column with a gradient of imidazole from 10 mM to 1 M. SDS-PAGE analysis, followed by Coomassie brilliant blue staining, showed that the fusion proteins were purified to near homogeneity.

### ChIP and genome-wide ChIP-chip

Cross-linked chromatin was prepared essentially as described previously [Kuras and Struhl 1999; Kuras et al. 2003]. Cells were grown exponentially to 0.6 OD<sub>600</sub> and cross-linked with 1% formaldehyde for 10 min. The 3HA- and 13Myc-tagged proteins were immunoprecipitated with 12CA5 and 9E10 antibodies, respectively; Pol II was immunoprecipitated with 8WG16 anti-CTD antibody (Covance), and bound to IgG magnetic beads (Dynabead). Immune complexes were washed as described previously [Kuras and Struhl 1999]. Cross-link reversal and DNA purification were performed as described [Kuras et al. 2003], except that the final elution was in 50 μL. Immunoprecipitated DNA was analyzed by quantitative real-time PCR on an ABI Prism 7000 or 7300 machine (Applied Biosystems). The PCR reactions were carried out in 25 μL containing 0.4 μM each primer, and 12.5 μL of mastermix SYBR green PCR reaction (Applied Biosystems). Relative quantification using a standard curve method was performed and the occupancy level for a specific fragment was defined as the ratio of immunoprecipitated DNA over total DNA. *GAL1* ORF region was used as a non-transcribed control.

Ligation-mediated PCR was done as described previously [Ren et al. 2000], except that amino-allyl conjugated dUTP (150 μM final) was used and only 30 cycles of PCR were performed. The PCR products were purified using a Microcon YM-30 filter. Amino-allyl modified DNA was recovered with 20 μL of H<sub>2</sub>O and the DNA was lyophilized. DNA was labeled as previously described [Harismendy et al. 2003], except that unincorporated dyes were removed using a QIAquick PCR Purification Kit (Qiagen). Labeled DNA was recovered with 100 μL of buffer EB, and ethanol precipitated. Hybridization and washing conditions were as described previously [Lee et al. 2006]. Microarrays were obtained from Agilent Technologies and feature 41,418 (G4486A) or 41,776 (G4493A) 60-mer oligonucleotide probe spots, with an average density of one probe each 266 bp of the yeast genome. Images of Cy5 and Cy3 fluorescence intensities were generated using the Genepix 4000B scanner (Molecular Devices) and extracted using GenePix Pro 6 software (Molecular Devices). At least two biological replicates were performed for each experiment.

### Data analysis

The computational data analysis was performed in R using the *limma* package [Smyth et al. 2005] from the bioconductor project (<http://www.bioconductor.org>). After subtraction of the local

background, the data from both channels were median normalized, and log ratios of signal intensities were generated for each feature. The log ratios were processed by fitting a linear model for each feature in order to calculate the average log ratio between replicates. *P*-values were then calculated by performing an empirical Bayes moderated *t*-statistic test, and adjusted for multiple testing by Benjamini and Hochberg false discovery rate (FDR) method. The complete raw data set is available at Array Express (<http://www.ebi.ac.uk/arrayexpress>) under accession number E-MTAB-10. Visualization of ChIP-enriched genomic regions was performed using an adaptation of the Ringo package [Toedling et al. 2007]. GO analysis was performed using GOstats package.

### Pol III transcriptome analysis

Microarray hybridization was performed as described previously [Ciesla et al. 2007]. Briefly, 20 μg of total RNA was reverse transcribed with specific primers designed to hybridize to the 3' end of all mature tRNAs. cDNA was labeled with Cy3 and Cy5 dyes (Amersham) and hybridized on a Pol III-specific microarray (O. Harismendy, pers. comm.) harboring all the different tRNA genes, as well as all the other genes transcribed by Pol III.

### In vitro transcription and primer extension assays

Standard Pol III in vitro transcription reactions on the *SUP4* template were performed as described previously [Ducrot et al. 2006; Alic et al. 2007] in 40 μL of transcription buffer. pRS316-*SUP4* plasmid (150 ng) was incubated for 20 min at 25°C in the presence of 100 ng of rTFIIIC, 20 ng of rTBP, 10 ng of rBdp1, 10 ng of rBrf1, 100 ng of highly purified Pol III, and 0.5 μg of purified B' fraction or different concentrations of TFIIS protein (wild-type or E291A mutant). Transcription was started by the addition of 600 μM A/C/GTP, 300 μM UTP, 25 μM [α-<sup>32</sup>P]UTP (400 Ci/mmol), and allowed to proceed for 15 min.

Primer extension assays were performed as described previously [Andrau and Werner 2001] on the *SUP4* template with 200 μM unlabeled ATP, CTP, UTP, and GTP. Single-round assays were performed as described, except that 125 μM unlabeled ATP, CTP, UTP, GTP, and 0.3 mg/mL heparin were used [Kasavetis and Steiner 2006].

The reaction products were analyzed by electrophoresis on denaturing polyacrylamide gels (7% for in vitro transcription and primer extension assay). Gels were autoradiographed using MR film with an intensifying screen (Kodak). Quantifications were performed with Image Quant software (Molecular Dynamics).

## Acknowledgments

We thank M. Wery, D. Després, E. Shematorova, M. Boguta, K. Struhl, D. Stillman, and O. Gadal for yeast strains or plasmids; C. Kane for anti-TFIIS antibody; the SPI (CEA/Saclay) for monoclonal antibodies; C. Conesa, O. Harismendy, and S.G.F. (CEA/Evry) for Pol III transcriptome arrays and protocols; M. Riva for the yeast Pol II preparation and advice on Pol II transcription assays; C. Carles, N. Ayoub, and N. Alic for the yeast Pol IIIΔ preparation and advice in cleavage assays; and C. Ducrot for Pol III transcription factor preparations. We also thank O. Lefebvre and A. Briand for useful advice, and A. Sentenac and C. Mann for critical reading of the manuscript. Y.G. and M.M. were supported by a "Contrat de Formation par la Recherche" from the CEA.

## References

Alic, N., Ayoub, N., Landrieux, E., Favry, E., Baudouin-Cornu, P., Riva, M., and Carles, C. 2007. Selectivity and proofread-

- ing both contribute significantly to the fidelity of RNA polymerase III transcription. *Proc. Natl. Acad. Sci.* **104**: 10400–10405.
- Andrau, J.C. and Werner, M. 2001. B<sup>+</sup>-associated factor(s) involved in RNA polymerase III preinitiation complex formation and start-site selection. *Eur. J. Biochem.* **268**: 5167–5175.
- Awrey, D.E., Weilbaecher, R.G., Hemming, S.A., Orlicky, S.M., Kane, C.M., and Edwards, A.M. 1997. Transcription elongation through DNA arrest sites. A multistep process involving both RNA polymerase II subunit RPB9 and TFIIS. *J. Biol. Chem.* **272**: 14747–14754.
- Awrey, D.E., Shimasaki, N., Koth, C., Weilbaecher, R., Olmsted, V., Kazanis, S., Shan, X., Arellano, J., Arrowsmith, C.H., Kane, C.M., et al. 1998. Yeast transcript elongation factor (TFIIS), structure and function. II: RNA polymerase binding, transcript cleavage, and read-through. *J. Biol. Chem.* **273**: 22595–22605.
- Borukhov, S., Sagitov, V., and Goldfarb, A. 1993. Transcript cleavage factors from *E. coli*. *Cell* **72**: 459–466.
- Braglia, P., Dugas, S.L., Donze, D., and Dieci, G. 2007. Requirement of Nhp6 proteins for transcription of a subset of tRNA genes and heterochromatin barrier function in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **27**: 1545–1557.
- Chedin, S., Riva, M., Schultz, P., Sentenac, A., and Carles, C. 1998. The RNA cleavage activity of RNA polymerase III is mediated by an essential TFIIS-like subunit and is important for transcription termination. *Genes & Dev.* **12**: 3857–3871.
- Chen, W., Tabor, S., and Struhl, K. 1987. Distinguishing between mechanisms of eukaryotic transcriptional activation with bacteriophage T7 RNA polymerase. *Cell* **50**: 1047–1055.
- Ciesla, M., Towpik, J., Graczyk, D., Oficjalska-Pham, D., Harismendy, O., Suleau, A., Balicki, K., Conesa, C., Lefebvre, O., and Boguta, M. 2007. Maf1 is involved in coupling carbon metabolism to RNA Polymerase III transcription. *Mol. Cell. Biol.* **27**: 7693–7702.
- Davis, C.A. and Ares Jr., M. 2006. Accumulation of unstable promoter-associated transcripts upon loss of the nuclear exosome subunit Rrp6p in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci.* **103**: 3262–3267.
- Ducrot, C., Lefebvre, O., Landrieux, E., Guirouilh-Barbat, J., Sentenac, A., and Acker, J. 2006. Reconstitution of the yeast RNA polymerase III transcription system with all recombinant factors. *J. Biol. Chem.* **281**: 11685–11692.
- Exinger, F. and Lacroute, F. 1992. 6-Azauracil inhibition of GTP biosynthesis in *Saccharomyces cerevisiae*. *Curr. Genet.* **22**: 9–11.
- Fish, R.N. and Kane, C.M. 2002. Promoting elongation with transcript cleavage stimulatory factors. *Biochim. Biophys. Acta* **1577**: 287–307.
- Geiduschek, E.P. and Kassavetis, G.A. 2001. The RNA polymerase III transcription apparatus. *J. Mol. Biol.* **310**: 1–26.
- Guglielmi, B., Soutourina, J., Esnault, C., and Werner, M. 2007. TFIIS elongation factor and Mediator act in conjunction during transcription initiation *in vivo*. *Proc. Natl. Acad. Sci.* **104**: 16062–16067.
- Harismendy, O., Gendrel, C.G., Soularue, P., Gidrol, X., Sentenac, A., Werner, M., and Lefebvre, O. 2003. Genome-wide location of yeast RNA polymerase III transcription machinery. *EMBO J.* **22**: 4738–4747.
- Hausner, W., Lange, U., and Musfeldt, M. 2000. Transcription factor S, a cleavage induction factor of the archaeal RNA polymerase. *J. Biol. Chem.* **275**: 12393–12399.
- Kassavetis, G.A. and Geiduschek, E.P. 2006. Transcription factor TFIIB and transcription by RNA polymerase III. *Biochem. Soc. Trans.* **34**: 1082–1087.
- Kassavetis, G.A. and Steiner, D.F. 2006. Nhp6 is a transcriptional initiation fidelity factor for RNA polymerase III transcription *in vitro* and *in vivo*. *J. Biol. Chem.* **281**: 7445–7451.
- Kettenberger, H., Armache, K.J., and Cramer, P. 2003. Architecture of the RNA polymerase II-TFIIS complex and implications for mRNA cleavage. *Cell* **114**: 347–357.
- Kim, B., Nesvizhskii, A.I., Rani, P.G., Hahn, S., Aebersold, R., and Ranish, J.A. 2007. The transcription elongation factor TFIIS is a component of RNA polymerase II preinitiation complexes. *Proc. Natl. Acad. Sci.* **104**: 16068–16073.
- Kuhn, C.D., Geiger, S.R., Baumli, S., Gartmann, M., Gerber, J., Jennebach, S., Mielke, T., Tschochner, H., Beckmann, R., and Cramer, P. 2007. Functional architecture of RNA polymerase I. *Cell* **131**: 1260–1272.
- Kulish, D. and Struhl, K. 2001. TFIIS enhances transcriptional elongation through an artificial arrest site *in vivo*. *Mol. Cell. Biol.* **21**: 4162–4168.
- Kuras, L. and Struhl, K. 1999. Binding of TBP to promoters *in vivo* is stimulated by activators and requires Pol II holoenzyme. *Nature* **399**: 609–613.
- Kuras, L., Borggreffe, T., and Kornberg, R.D. 2003. Association of the Mediator complex with enhancers of active genes. *Proc. Natl. Acad. Sci.* **100**: 13887–13891.
- Landrieux, E., Alic, N., Ducrot, C., Acker, J., Riva, M., and Carles, C. 2006. A subcomplex of RNA polymerase III subunits involved in transcription termination and reinitiation. *EMBO J.* **25**: 118–128.
- Lee, T.I., Johnstone, S.E., and Young, R.A. 2006. Chromatin immunoprecipitation and microarray-based analysis of protein location. *Nat. Protoc.* **1**: 729–748.
- Malagon, F., Tong, A.H., Shafer, B.K., and Strathern, J.N. 2004. Genetic interactions of *DST1* in *Saccharomyces cerevisiae* suggest a role of TFIIS in the initiation–elongation transition. *Genetics* **166**: 1215–1227.
- Martens, J.A., Laprade, L., and Winston, F. 2004. Intergenic transcription is required to repress the *Saccharomyces cerevisiae* *SER3* gene. *Nature* **429**: 571–574.
- Mason, P.B. and Struhl, K. 2005. Distinction and relationship between elongation rate and processivity of RNA polymerase II *in vivo*. *Mol. Cell* **17**: 831–840.
- Moqtaderi, Z. and Struhl, K. 2004. Genome-wide occupancy profile of the RNA polymerase III machinery in *Saccharomyces cerevisiae* reveals loci with incomplete transcription complexes. *Mol. Cell. Biol.* **24**: 4118–4127.
- Nakanishi, T., Shimoaraiso, M., Kubo, T., and Natori, S. 1995. Structure–function relationship of yeast S-II in terms of stimulation of RNA polymerase II, arrest relief, and suppression of 6-azauracil sensitivity. *J. Biol. Chem.* **270**: 8991–8995.
- Natori, S., Takeuchi, K., Takahashi, K., and Mizuno, D. 1973. DNA dependent RNA polymerase from Ehrlich ascites tumor cells. II. Factors stimulating the activity of RNA polymerase II. *J. Biochem.* **73**: 879–888.
- Nonet, M., Scafe, C., Sexton, J., and Young, R. 1987. Eucaryotic RNA polymerase conditional mutant that rapidly ceases mRNA synthesis. *Mol. Cell. Biol.* **7**: 1602–1611.
- Olmsted, V.K., Awrey, D.E., Koth, C., Shan, X., Morin, P.E., Kazanis, S., Edwards, A.M., and Arrowsmith, C.H. 1998. Yeast transcript elongation factor (TFIIS), structure and function. I: NMR structural analysis of the minimal transcriptionally active region. *J. Biol. Chem.* **273**: 22589–22594.
- Opalka, N., Chlenov, M., Chacon, P., Rice, W.J., Wriggers, W., and Darst, S.A. 2003. Structure and function of the transcription elongation factor GreB bound to bacterial RNA polymerase. *Cell* **114**: 335–345.
- Pokholok, D.K., Hannett, N.M., and Young, R.A. 2002. Exchange of RNA polymerase II initiation and elongation factors during gene expression *in vivo*. *Mol. Cell* **9**: 799–809.

- Prather, D.M., Larschan, E., and Winston, F. 2005. Evidence that the elongation factor TFIIS plays a role in transcription initiation at *GAL1* in *Saccharomyces cerevisiae*. *Mol. Cell Biol.* **25**: 2650–2659.
- Ren, B., Robert, F., Wyrick, J.J., Aparicio, O., Jennings, E.G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E., et al. 2000. Genome-wide location and function of DNA binding proteins. *Science* **290**: 2306–2309.
- Roberts, D.N., Stewart, A.J., Huff, J.T., and Cairns, B.R. 2003. The RNA polymerase III transcriptome revealed by genome-wide localization and activity–occupancy relationships. *Proc. Natl. Acad. Sci.* **100**: 14695–14700.
- Roberts, D.N., Wilson, B., Huff, J.T., Stewart, A.J., and Cairns, B.R. 2006. Dephosphorylation and genome-wide association of Maf1 with Pol III-transcribed genes during repression. *Mol. Cell* **22**: 633–644.
- Sawadogo, M., Huet, J., and Fromageot, P. 1980a. Similar binding site for P37 factor on yeast RNA polymerases A and B. *Biochem. Biophys. Res. Commun.* **96**: 258–264.
- Sawadogo, M., Sentenac, A., and Fromageot, P. 1980b. Interaction of a new polypeptide with yeast RNA polymerase B. *J. Biol. Chem.* **255**: 12–15.
- Sawadogo, M., Lescure, B., Sentenac, A., and Fromageot, P. 1981. Native deoxyribonucleic acid transcription by yeast RNA polymerase–P37 complex. *Biochemistry* **20**: 3542–3547.
- Schnapp, G., Graveley, B.R., and Grummt, I. 1996. TFIIS binds to mouse RNA polymerase I and stimulates transcript elongation and hydrolytic cleavage of nascent rRNA. *Mol. Gen. Genet.* **252**: 412–419.
- Shilatifard, A., Conaway, R.C., and Conaway, J.W. 2003. The RNA polymerase II elongation complex. *Annu. Rev. Biochem.* **72**: 693–715.
- Smyth, G.K., Michaud, J., and Scott, H.S. 2005. Use of within-array replicate spots for assessing differential expression in microarray experiments. *Bioinformatics* **21**: 2067–2075.
- Toedling, J., Sklyar, O., and Huber, W. 2007. Ringo—An R/Bioconductor package for analyzing ChIP–chip readouts. *BMC Bioinformatics* **8**: 221; doi: 10.1186/1471-2105-8-221.
- Ubukata, T., Shimizu, T., Adachi, N., Sekimizu, K., and Nakanishi, T. 2003. Cleavage, but not read-through, stimulation activity is responsible for three biologic functions of transcription elongation factor S-II. *J. Biol. Chem.* **278**: 8580–8585.
- Wery, M., Shematorova, E., Van Driessche, B., Vandenhoute, J., Thuriaux, P., and Van Mullem, V. 2004. Members of the SAGA and Mediator complexes are partners of the transcription elongation factor TFIIS. *EMBO J.* **23**: 4232–4242.
- Wyers, F., Rougemaille, M., Badis, G., Rousselle, J.C., Dufour, M.E., Boulay, J., Regnault, B., Devaux, F., Namane, A., Seraphin, B., et al. 2005. Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. *Cell* **121**: 725–737.
- Zaros, C. and Thuriaux, P. 2005. Rpc25, a conserved RNA polymerase III subunit, is critical for transcription initiation. *Mol. Microbiol.* **55**: 104–114.

## SUPPLEMENTARY DATA

### Supplementary Materials and methods

#### *Yeast strains and plasmids*

All *S. cerevisiae* strains are described in Table S2. Except for MB159-2C, they all derive from the YPH499/YPH500 standard strains of Sikorski and Hieter (1989). The 3HA-C160 (MW671) strain has been described elsewhere (Dieci et al. 1995). The 3HA-TFIIS (YGH2) strain was derived from YCO41 strain (a gift from M. Wery). *URA3* and *ADE2* cassettes were isolated by BglII restriction of plasmids pFL44L and pASZ10, respectively. Correct integration of the *URA3* and *ADE2* cassettes was verified by PCR analysis and sequencing. Correct expression of the tagged protein was verified by western blotting. The ESH1 (*dst1-Δ*) strain (Wery et al. 2004) was transformed by pRS425-pPGK, pRS425-DST1, pRS425-dst1R200A or pRS425-dst1E291A plasmids. Strains YGH10 (*SUP11 dst1-Δ*), YGH9 (*dst1-Δ*) and YGH16 (*nhp6a-Δ nhp6b-Δ dst1-Δ*) were constructed by deleting the chromosomal *DST1* copy with a *dst1Δ::KanMX4* cassette (Euroscarf: <http://www.uni-frankfurt.de/fb15/mikro/euroscarf/>) in MB159-2c strain (a gift from M. Boguta) bearing *SUP11* gene, YGH2 strain or DY2381 strain (a gift from D. Stillman), respectively. The *rpb1-1* (D788-4a) and *rpc25-S100P* (D792-3a) strains were constructed by standard meiotic crosses and transformed by Yep351 and pYeB-B220 or pRS315 and pRS315-RPC25 plasmids, respectively. The 3HA-TFIIS RPC160-13Myc (YGH11) and RPC160-13Myc *dst1-Δ* (YGH12) strains were constructed by insertion of 13Myc epitopes followed by a *His3MX6* marker in YGH2 or YGH9 strains, respectively, using the standard one-step method (Longtine et al. 1998). The BDP1-3HA (MW4035) and TFC1-13Myc (yOH1) strains have been described previously (Harismendy et al. 2003). Their chromosomal *DST1* copy was deleted using the standard one-step method to obtain YGH13 and YGH14 strains,



respectively (Longtine et al. 1998). RPC11-3HA (YGH15) strain was constructed by insertion of 3HA epitopes followed by a *His3MX6* marker in YPH500 strain transformed with pRS316-RPC11 plasmid carrying wild-type Rpc11 (Longtine et al. 1998). The strains KY1239 and KY1152 (a gift from K. Struhl) contain the T7 promoter driving *HIS3* structural gene transcription and express or not the T7 RNA polymerase on a *URA3* plasmid (Chen et al. 1987).

All plasmids are described in Table S3. pDEST17-DST1, pDEST17-dst1E291A and pRS425-DST1 plasmids were constructed using the GATEWAY™ Invitrogen cloning method (Walhout et al. 2000). The sequences of interest were transferred from the entry vectors pDonr-DST1 (Guglielmi et al. 2007) and pDonr-dst1E291A into pDEST17 or pRS425-pPGK vector by LR reaction. The sequence of the oligonucleotides used in this study can be found in Table S4.

### ***Northern blotting***

Cells were grown overnight, diluted to 0.2 OD<sub>600</sub>, and exponentially grown in selective medium at 30°C until they reached 0.6 OD<sub>600</sub> or shifted at 16°C for 8h. Total RNA was extracted as described previously (Schmitt et al. 1990). Small size RNAs were separated by electrophoresis on a denaturing 8% polyacrylamide gel run in TBE buffer. The gel was then blotted onto a positively charged Nylon membrane (Roche) using the bio-Rad Trans-Blot cell. As previously described (Harismendy et al. 2003), a PCR product was labeled by nick translation with [ $\alpha$ -<sup>32</sup>P]dCTP and used as a probe. Signal analysis and quantification were performed using Phosphorimager and Image Quant software (Molecular Dynamics).

### ***Quantitative RT-PCR.***

Total RNA was extracted as described previously (Schmitt et al. 1990). Reverse transcription of 5- $\mu$ g RNA samples was performed in the presence of SuperScript II and random primers (Invitrogen) for 1 h at 42°C in the appropriate buffer. 25S rRNA was used as internal control for normalization.

### ***Coimmunoprecipitations***

Noncrosslinked whole cells extract preparation, immunoprecipitations and Western blotting analysis were performed as described previously (Soutourina et al. 2006). When crosslinked cells were used, extract preparations and immunoprecipitations were as described for chromatin immunoprecipitation, except that samples were eluted directly into SDS sample buffer and incubated for 15 min at 95°C before running on the gel.

### ***Cleavage and elongation assays***

To analyze the cleavage and elongation kinetics, a 17-mer assay was performed as described previously (Chedin et al. 1998). Stable ternary complexes were formed in 100  $\mu$ l of transcription buffer in the absence of GTP and with 3  $\mu$ M [ $\alpha$ -<sup>32</sup>P]UTP (400 Ci/mmol). The complexes were purified by size-exclusion chromatography on 1 ml CL-2B columns as described previously (Steinberg and Burgess 1992; Dieci et al. 1995). Transcript elongation was analyzed by resuming elongation by 17 nt halted ternary complexes by the addition of ATP, CTP, GTP (600  $\mu$ M) and UTP (30  $\mu$ M). Reactions were stopped at regular intervals and the reaction products were analysed on 15% denaturing polyacrylamide gels. Cleavage reactions were initiated by the addition to the labeled ternary complexes of transcription buffer containing MgCl<sub>2</sub> to bring the final concentration to 5 mM. Cleavage products were analyzed as described for elongation reactions. Gels were autoradiographed using MR film

with an intensifying screen (Kodak). Quantifications were performed with Image Quant software (Molecular Dynamics).



## Supplementary Tables

**Table S1. Overrepresented GO biological process categories in Pol II and TFIIIS enriched genes.**

	Pol II			TFIIIS		
	GO ID	GO Term	p-value	GO ID	GO Term	p-value
1	GO:0009987	cellular process	1.12E-56	GO:0006412	translation	2.01E-46
2	GO:0042254	ribosome biogenesis and assembly	6.53E-40	GO:0009058	biosynthetic process	6.87E-37
3	GO:0016043	cell organization and biogenesis	1.03E-37	GO:0009987	cellular process	1.88E-36
4	GO:0044238	primary metabolic process	7.35E-37	GO:0044249	cellular biosynthetic process	1.95E-36
5	GO:0044237	cellular metabolic process	5.53E-36	GO:0009059	macromolecule biosynthetic process	6.63E-32
6	GO:0022613	ribonucleoprotein complex biogenesis and assembly	1.52E-34	GO:0042254	ribosome biogenesis and assembly	6.32E-25
7	GO:0008152	metabolic process	1.91E-34	GO:0022613	ribonucleoprotein complex biogenesis and assembly	1.79E-23
8	GO:0043170	macromolecule metabolic process organelle	6.37E-28	GO:0044237	cellular metabolic process	1.09E-20
9	GO:0006996	organization and biogenesis	1.09E-27	GO:0008152	metabolic process	1.10E-20
10	GO:0009058	biosynthetic process	8.21E-26	GO:0044238	primary metabolic process cellular	6.41E-20
11	GO:0006412	translation	2.06E-25	GO:0044260	macromolecule metabolic process	2.25E-13
12	GO:0044260	cellular macromolecule metabolic process	4.54E-25	GO:0044267	cellular protein metabolic process	2.82E-13
13	GO:0044267	cellular protein metabolic process	7.03E-25	GO:0019538	protein metabolic process	1.47E-12
14	GO:0009059	macromolecule biosynthetic process	2.91E-24	GO:0016043	cell organization and biogenesis	3.26E-11
15	GO:0006364	rRNA processing	3.18E-24	GO:0006364	rRNA processing	3.32E-11

**Table S2. Yeast strains**

<b>Name</b>	<b>Tagged protein</b>	<b>Genotype</b>	<b>Reference</b>
MW671	3HA-Rpc160	<i>MATa ade2-1 lys2-801 ura3-52 trp1-Δ63 his3-Δ200 leu2-Δ1 rpc160-Δ1::HIS3 pC160-240 (TRP1 CEN4 RPC160-240)</i>	Dieci et al. 1995
YCO41	3HA-TFIIS	<i>MATa ade2-1 lys2-801 ura3-52 trp1-Δ63 his3-Δ200 leu2-Δ1 TRP1::pADH::3HA::dst1</i>	M. Wery, unpublished
YGH2	3HA-TFIIS	<i>MATa lys2-801 trp1-Δ63 his3-Δ200 leu2-Δ1 TRP1::pADH::3HA::dst1</i>	This work
ESH1		<i>MATa ade2-1 lys2-801 ura3-52 trp1-Δ63 his3-Δ200 leu2-Δ1 dst1::hisG</i>	Wery et al. 2004
YGH9		<i>MATa lys2-801 trp1-Δ63 his3-Δ200 leu2-Δ1 dst1::KanMX4</i>	This work
MB159-2c		<i>MATα SUP11 ade2-1 ura3-52 trp1-Δ63 leu2-Δ1</i>	M. Boguta
YGH10		<i>MATα SUP11 ade2-1 ura3-52 trp1-Δ63 leu2-Δ1 dst1::KanMX4</i>	This work
D788-4a	3HA-TFIIS Rpc160-13Myc	<i>MATα trp1-Δ63 his3-Δ200 leu2-Δ1 rpb1-1 RPC160::13Myc::KanMX6 TRP1::pADH::3HA::dst1</i>	This work
D792-3a	3HA-TFIIS Rpc160-13Myc	<i>MATa trp1-Δ63 his3-Δ200 leu2-Δ1 rpc25-S100P RPC160::13Myc::KanMX6 TRP1::pADH::3HA::dst1</i>	This work
YGH11	3HA-TFIIS Rpc160-13Myc	<i>MATa lys2-801 trp1-Δ63 his3-Δ200 leu2-Δ1 TRP1::pADH::3HA::dst1 RPC160::13Myc::His3MX6</i>	This work
YGH12	Rpc160-13Myc	<i>MATa lys2-801 trp1-Δ63 his3-Δ200 leu2-Δ1 dst1::KanMX4 RPC160::13Myc::His3MX6</i>	This work
MW4035	Bdp1-3HA	<i>MATa ade2-1 lys2-801 ura3-52 trp1-Δ63 his3-Δ200 leu2-Δ1 BDP1-3HA::KanMX6</i>	Harismendy et al. 2003
yOH1	Tfc1-13Myc	<i>MATα ade2-1 lys2-801 ura3-52 trp1-Δ63 his3-Δ200 leu2-Δ1 TFC1-13Myc::KanMX6</i>	Harismendy et al. 2003
YGH13	Bdp1-3HA	<i>MATa ade2-1 lys2-801 ura3-52 trp1-Δ63 his3-Δ200 leu2-Δ1 BDP1-3HA::KanMX6 dst1::His3MX6</i>	This work
YGH14	Tfc1-13Myc	<i>MATα ade2-1 lys2-801 ura3-52 trp1-Δ63 his3-Δ200 leu2-Δ1 TFC1-13Myc::KanMX6 dst1::His3MX6</i>	This work
YGH15	Rpc11-3HA	<i>MATα ade2-1 lys2-801 ura3-52 trp1-Δ63 his3-Δ200 leu2-Δ1 RPC11-3HA::His3MX6 pRS316-RPC11 (URA3 CEN RPC11)</i>	This work
DY2381		<i>MATα ade2 can1 his3 leu2 trp1 ura3 nhp6a:URA3 nhp6b:HIS</i>	Eriksson et al. 2004
YGH16		<i>MATα ade2 can1 his3 leu2 trp1 ura3 nhp6a::URA3 nhp6b::HIS dst1::KanMX4</i>	This work

**Table S3. Plasmids**

<b>Name</b>	<b>Genotype</b>	<b>Reference</b>
pDEST17	<i>Amp</i>	Invitrogen
pDEST17-DST1	<i>Amp DST1</i>	This work
pDEST17-dst1E291A	<i>Amp dst1E291A</i>	This work
pDonr-DST1	<i>Km attL1-DST1-attL2</i>	Guglielmi et al. 2007
pDonr-dst1E291A	<i>Km attL1-dst1E291A-attL2</i>	D. Després
Yep351	<i>Amp 2μ LEU2</i>	Hill et al. 1986
pYeB-B220	<i>Amp 2μ LEU2 RPB1</i>	Voutsina et al. 1999
pRS315	<i>Amp CEN LEU2</i>	Sikorski and Hieter 1989
pRS315-RPC25	<i>Amp CEN LEU2 RPC25</i>	O. Gadai
pFL44L	<i>Amp 2μ URA3</i>	Bonneaud et al. 1991
pASZ10	<i>Amp CEN ADE2</i>	Stotz and Linder 1990
pRS425-pPGK	<i>Amp 2μ LEU2</i>	O. Gadai
pRS425-DST1	<i>Amp 2μ LEU2 DST1</i>	This work
pRS425-dst1R200A	<i>Amp 2μ LEU2 dst1R200A</i>	D. Després
pRS425-dst1E291A	<i>Amp 2μ LEU2 dst1E291A</i>	D. Després
pRS316-RPC11	<i>Amp CEN URA3 RPC11</i>	Chedin et al. 1998

**Table S4. Oligonucleotides**

<b>Name</b>	<b>Forward</b>	<b>Reverse</b>
<i>ADH1</i>	GGGTATTGACGGTGGTGAAG	AAACGTTGATGACACCGTGA
<i>RPL28</i>	TCGGTAAGCACAGAAAGCAC	GGCTTCCAGAAATGAGCTTG
<i>RPL17a</i>	ATTCAAAATGCGCGGAAAAT	TCAGCTTGACAGTTCATTCCA
<i>ACT1</i>	CATGTTCCCAGGTATTGCCGA	GTCAAAGAAGCCAAGATAGAA
<i>GAL1</i>	AAAGAAACTTGCACCGGAAA	GGCCCATATTGCTTTAACA
<i>SCR1-1</i>	CCAGGACATCCATAGCTTGTG	ATGAAAAGTTCCTGGCGATG
<i>SCR1-2</i>	CATCGCCAGGAACCTTTTCAT	ACAGCCTAGCACAATTGGAA
<i>SCR1-3</i>	CTTTCTGGTGGGATGGGATA	TTTACGACGGAGGAAAGACG
<i>SCR1-4</i>	GTCCTGGGCAGAGCTGTCT	AAGGTGGAGCCCTAAGGA
<i>SCR1-5</i>	ACCGCTGTTAGGGGAGTTTT	CCAAATTAACCGCCGAAG
<i>SCR1-6</i>	CGGTGCCATCAGGATTTACT	CTTCCAACATCCCTCATTGG
<i>SCR1-7</i>	TTTTCGAATATAAATGACGATTGG	TGTCGCTACTCACTCTACAACCA
<i>SNR6</i>	GTCATCTTCTGGACCTCATGTGA	AGGGGAAGTGTGATCATCTCTGT
<i>RPR1</i>	ATGGTACGCTGTGGTGCTC	CCATAGGTGGGGATCCTTCT
<i>tDNA</i> <i>Phe</i>	GACGCTTGGACCATTATATAAAGCAC	CCATAAGAGAAGGAGCAGTCAAGTTCA
<i>tDNA</i> <i>Met</i>	GCTTCAGTAGCTCAGTAGGAA	TGCTCCAGGGGAGGTTTC
<i>tDNA</i> <i>Leu</i>	GGGTTCGAATCCCTTAGCTC	GATGGCGCCTTGACTAACTT
<i>rDNA1</i>	TCGCGTATGGTCACCCACTAC	GTTGCGGCCATATCTACCAGA
<i>rDNA2</i>	AAGATTGCAGCACCTGAGTTTCG	ATGTCTGGACCCTGCCCTCATAT
<i>rDNA3</i>	CCGTTATTGGTAGGAGTGTGGTG	ACGGAAATACGCTTCAGAGACC
<i>rDNA4</i>	GTGTGAGGAAAAGTAGTTGGGAGGTA	GACGAGGCCATTTACAAAAACATAAC
<i>rDNA5</i>	GCGAACCAGGACTTTTACTTTGA	AACAAAATAGAACCAAACGTCTATTC
<i>rDNA6</i>	GGTCTGTGATGCCCTTAGACG	AGTTTCACAAGATTACCAAGACCTCTC
<i>rDNA7</i>	AATATTAAAAACTTTCAACAACGGATCTCT	CGATGATTCACGGAATTCTGC
<i>rDNA8</i>	GGTGGTAAATTCCATCTAAAGCTAAATATT	CACGTACTTTTCACTCTCTTTTCAAA
<i>rDNA9</i>	TATGAGGTAAAGCGAATGATTAGAGG	CACGTTCAATTAAGTAACAAGGACTTCT
<i>rDNA10</i>	GGAGGAGTTATCTTTTCTTCTTAACAGCT	AAGGTGCTGGCCTCTTCCA
<i>rDNA11</i>	AAAGAAGACCCTGTTGAGCTTGA	GTATTTCACTGGCGCCGAA
<i>rDNA12</i>	ATCATTTGTATACGACTTAGATGTACAACG	AACAAATCAGACAACAAAGGCTTAATC
<i>rDNA13</i>	CTCATTTCTATAGTTAACAGGACATGC	TTCATTGTCTCTTACATCTTTCTTGG
<i>rDNA14</i>	GTACATATCAAGTAGTAGCAACCCAATGAG	ACCATTGATTTCAGAAAAATTTCG
<i>HIS3-1</i>	GAGCAGAAAGCCCTAGTAAAGC	AATCACTTGCGATTGTGTGG
<i>HIS3-2</i>	ACGACCATCACACCACTGAA	AGATCTACCACCGCTCTGGA
<i>HIS3-3</i>	GATGATCCCGCATTTTCTTG	GCGAGGTGGCTTCTCTTATG
<i>dw HIS3</i>	TGCATCATTCTATACGTGTCATTC	ACCACTTGCCACCTATCACC
<i>DED1-1</i>	ATAACAACAACGGCGGCTAC	GCCATCGATCCATCTACCAC
<i>DED1-2</i>	GGCCCGTGATTTCTTAAGTG	ACCGTCAGTGGATGCAGAC
<i>DED1-3</i>	GTAACACCGGTCTTGCCACT	GGCTGTTGCTTCTGCTACCT

## Supplementary Figure Legends

### Figure S1. ChIP analysis of TFIIS enrichment on T7 RNA polymerase-transcribed *HIS3* gene.

(A) Schematic organization of *HIS3-DED1* locus. The location of the PCR fragments amplified in ChIP analyses are indicated by black lines. The arrows indicate the transcription start site from the T7 and *DED1* promoters.

(B) Quantitative ChIP analysis of TFIIS enrichment on T7 RNA polymerase-transcribed *HIS3* gene or Pol II-transcribed *DED1* gene. The strains containing the T7 promoter for *HIS3* structural gene transcription and expressing (left panel) or not (right panel) the T7 RNA polymerase from *URA3* plasmid were grown in Casamino acids medium supplemented with adenine and tryptophan (CAW). Immunoprecipitations were performed using antibodies against TFIIS. The *GALI* ORF was used as a control. Error bars represent the standard deviation between at least three biological replicates. The background level is indicated by a dotted line.

(C) The transcription of *HIS3* gene by T7 RNA polymerase was measured by quantitative RT-PCR. 25S rRNA was used as an internal control for normalization.

### Figure S2. Example of enrichment profiles of Pol II and TFIIS on Pol II-transcribed genes.

The genomic positions of the probes are indicated along the x axis and represented by black points or circles. Watson strand-transcribed genes are colored in blue and Crick strand-transcribed genes are colored in green. The enrichment ratio is reported along the y axis.

(A, B) Enrichment profile of Pol II (blue) and TFIIS (red) on *ADHI* and *PGKI* genes from genome-wide ChIP in YPD medium.

(C, D) Enrichment profile of Pol II (blue) and TFIIS (red) from genome-wide ChIP in YPD medium on intergenic regions of chromosome V encoding short unstable RNAs (colored in violet) synthesized by Pol II.

**Figure S3. Genome-wide location analysis of TFIIS under NTP-depletion.**

(A) Effect of MPA on the genome-wide distribution of TFIIS. The genome-wide enrichment of TFIIS in the absence or presence of MPA (10 µg/ml, 4h) was assessed from ChIP-chip experiments. The dotted line corresponds to  $y=x$ .

(B, C) Enrichment profile of TFIIS in various conditions on *PYK1* and *ILV5* genes. The enrichment of TFIIS in rich medium (blue), minimal medium (green) and in the presence of MPA (violet) was assessed from ChIP-chip experiments. The genomic positions of the probes on chromosome 1 or 15 are reported along the x axis and represented by black points or circles. Watson strand-transcribed gene *PYK1* is colored in blue and Crick strand-transcribed gene *ILV5* is colored in green. The enrichment ratio is reported along the y axis.

**Figure S4. Example of enrichment profiles of TFIIS on Pol III-transcribed genes.**

Enrichment profile of TFIIS or Pol III from genome-wide ChIP in YPD medium on *SNR6* (A), *RPR1* (B) and *tDNAGly* (C) genes. The genomic positions of probe regions are indicated along the x axis and represented by black points as in Fig. 2B. Watson strand-transcribed genes are colored in blue and Crick strand-transcribed genes are colored in green. The enrichment ratio is reported along the y axis.

**Figure S5. Coimmunoprecipitation of TFIIS with Pol III transcription machinery.**

(A) TFIIS does not coimmunoprecipitate with Pol III, TFIIC and TFIIB before formaldehyde crosslinking. Bdp1-HA was immunoprecipitated from crude extracts with anti-

HA and Rpc160-Myc or Tfc1-Myc with anti-Myc antibodies. The immunoprecipitated proteins were revealed with anti-HA (Bdp1-HA), anti-Myc (Rpc160-Myc or Tfc1-Myc) or anti-TFIIS antibodies. The presence of HA- or Myc-tagged alleles is indicated above each lane. The position of the proteins is indicated on the side of the Western blots. The amount of proteins in crude extracts (Input) was also analyzed by Western blotting.

(B) TFIIS coimmunoprecipitates with Pol III and TFIIC after formaldehyde crosslinking. The extracts from formaldehyde-crosslinked cells were used to CoIP experiments. Bdp1-HA and Rpc11-HA was immunoprecipitated from extracts with anti-HA and Rpc160-Myc or Tfc1-Myc with anti-Myc antibodies.

**Figure S6. Example of enrichment profiles of Pol II and Pol III in wild-type or *dst1-Δ* mutant at 16°C.**

The genomic positions of probe regions are indicated along the x axis and represented by black points. Watson strand-transcribed genes are colored in blue and Crick strand-transcribed genes are colored in green. The enrichment ratio is reported along the y axis.

(A, B) Enrichment profile of Pol II on *PYK1* and *ADH1* genes from genome-wide ChIP. Wild-type strain (blue) and *dst1-Δ* mutant (violet) were grown in YPD medium and shifted to 16°C for 8h.

(C, D) Enrichment profile of Pol III on *SCR1* and *tDNAGly* genes from genome-wide ChIP. Wild-type strain (blue) and *dst1-Δ* mutant (violet) were grown in YPD medium and shifted to 16°C for 8h.

**Figure S7. Primer extension analysis of single round transcription *in vitro* and transcription *in vivo*.**

(A) Primer extension analysis of start sites used by Pol III on the *SUP4* gene *in vitro*. Single round transcriptions were performed in the absence of TFIIS, in the presence of wild-type TFIIS, mutant TFIIS-E291A or with B" fraction. Single round transcription and primer extensions were performed as described in Materials and methods using a probe hybridizing within the *SUP4* intron. Positions of the major transcription start sites are indicated.

(B) Primer extension analysis of start sites used by Pol III on the *SUP4* gene *in vivo*. Total RNAs from *dst1-Δ nhp6a-Δ nhp6b-Δ* and *nhp6a-Δ nhp6b-Δ* strains were analyzed by primer extension as described in Materials and methods.

**Figure S8. Time-course analysis of transcription elongation by Pol III.**

Labeled Pol III ternary complexes stalled at the 17<sup>th</sup> nucleotide were isolated and then incubated with unlabeled nucleotides in the absence or presence of TFIIS for various periods of time.

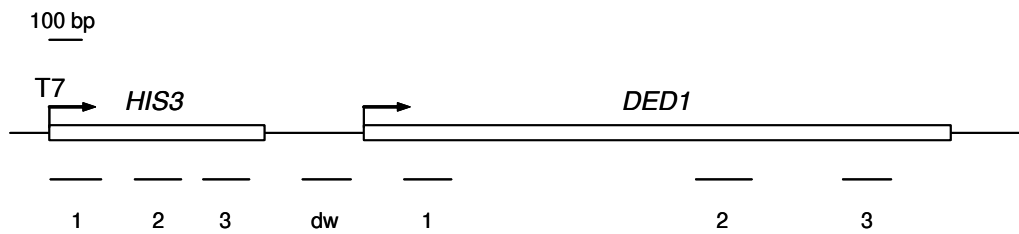


## Supplementary References

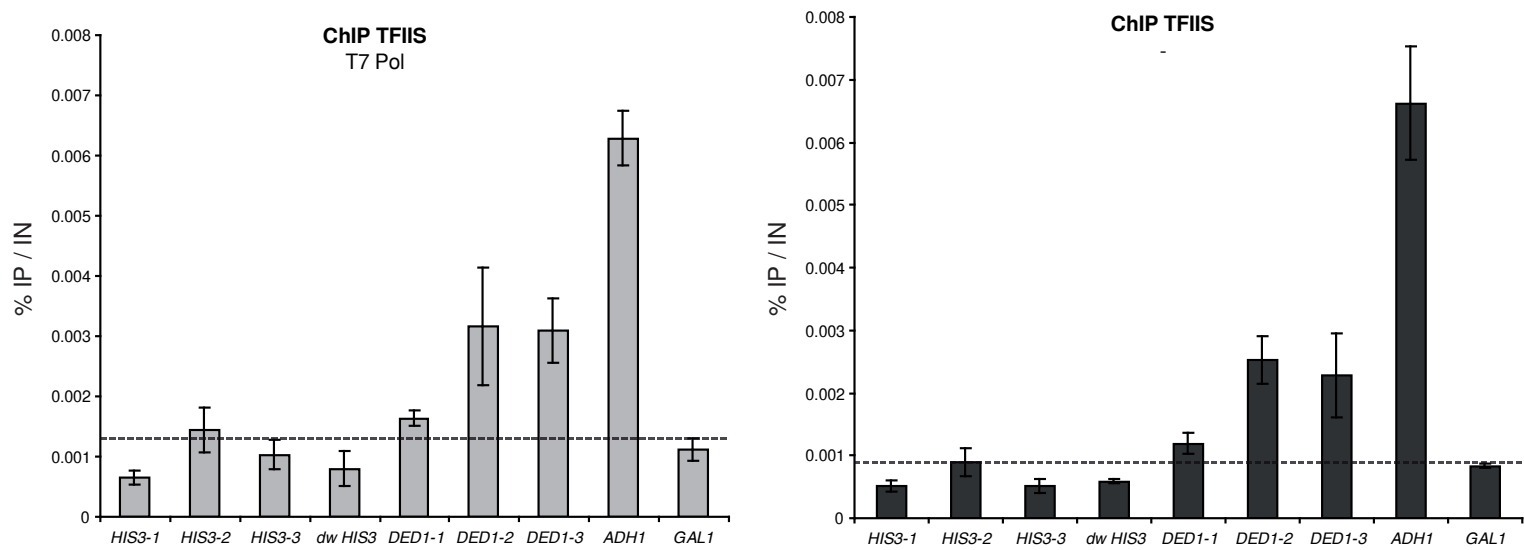
- Bonneaud, N., Ozier-Kalogeropoulos, O., Li, G.Y., Labouesse, M., Minvielle-Sebastia, L., and Lacroute, F. 1991. A family of low and high copy replicative, integrative and single-stranded *S. cerevisiae*/*E. coli* shuttle vectors. *Yeast* **7**(6): 609-615.
- Chedin, S., Riva, M., Schultz, P., Sentenac, A., and Carles, C. 1998. The RNA cleavage activity of RNA polymerase III is mediated by an essential TFIIS-like subunit and is important for transcription termination. *Genes Dev* **12**(24): 3857-3871.
- Chen, W., Tabor, S., and Struhl, K. 1987. Distinguishing between mechanisms of eukaryotic transcriptional activation with bacteriophage T7 RNA polymerase. *Cell* **50**(7): 1047-1055.
- Dieci, G., Hermann-Le Denmat, S., Lukhtanov, E., Thuriaux, P., Werner, M., and Sentenac, A. 1995. A universally conserved region of the largest subunit participates in the active site of RNA polymerase III. *Embo J* **14**(15): 3766-3776.
- Eriksson, P., Biswas, D., Yu, Y., Stewart, J.M., and Stillman, D.J. 2004. TATA-binding protein mutants that are lethal in the absence of the Nhp6 high-mobility-group protein. *Mol Cell Biol* **24**(14): 6419-6429.
- Guglielmi, B., Soutourina, J., Esnault, C., and Werner, M. 2007. TFIIS elongation factor and Mediator act in conjunction during transcription initiation *in vivo*. *Proc Natl Acad Sci U S A* **104**(41): 16062-16067.
- Harismendy, O., Gendrel, C.G., Soularue, P., Gidrol, X., Sentenac, A., Werner, M., and Lefebvre, O. 2003. Genome-wide location of yeast RNA polymerase III transcription machinery. *Embo J* **22**(18): 4738-4747.
- Hill, J.E., Myers, A.M., Koerner, T.J., and Tzagoloff, A. 1986. Yeast/*E. coli* shuttle vectors with multiple unique restriction sites. *Yeast* **2**(3): 163-167.
- Longtine, M.S., McKenzie, A., 3rd, Demarini, D.J., Shah, N.G., Wach, A., Brachat, A., Philippsen, P., and Pringle, J.R. 1998. Additional modules for versatile and economical PCR-based gene deletion and modification in *Saccharomyces cerevisiae*. *Yeast* **14**(10): 953-961.
- Schmitt, M.E., Brown, T.A., and Trumpower, B.L. 1990. A rapid and simple method for preparation of RNA from *Saccharomyces cerevisiae*. *Nucleic Acids Res* **18**(10): 3091-3092.
- Sikorski, R.S. and Hieter, P. 1989. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. *Genetics* **122**(1): 19-27.
- Soutourina, J., Bordas-Le Floch, V., Gendrel, G., Flores, A., Ducrot, C., Dumay-Odelot, H., Soularue, P., Navarro, F., Cairns, B.R., Lefebvre, O., and Werner, M. 2006. Rsc4 connects the chromatin remodeler RSC to RNA polymerases. *Mol Cell Biol* **26**(13): 4920-4933.
- Steinberg, T.H. and Burgess, R.R. 1992. Tagetitoxin inhibition of RNA polymerase III transcription results from enhanced pausing at discrete sites and is template-dependent. *J Biol Chem* **267**(28): 20204-20211.
- Stotz, A. and Linder, P. 1990. The *ADE2* gene from *Saccharomyces cerevisiae*: sequence and new vectors. *Gene* **95**(1): 91-98.
- Voutsina, A., Riva, M., Carles, C., and Alexandraki, D. 1999. Sequence divergence of the RNA polymerase shared subunit ABC14.5 (Rpb8) selectively affects RNA polymerase III assembly in *Saccharomyces cerevisiae*. *Nucleic Acids Res* **27**(4): 1047-1055.

- Walhout, A.J., Temple, G.F., Brasch, M.A., Hartley, J.L., Lorson, M.A., van den Heuvel, S., and Vidal, M. 2000. GATEWAY recombinational cloning: application to the cloning of large numbers of open reading frames or ORFeomes. *Methods Enzymol* **328**: 575-592.
- Wery, M., Shematorova, E., Van Driessche, B., Vandenhoute, J., Thuriaux, P., and Van Mullem, V. 2004. Members of the SAGA and Mediator complexes are partners of the transcription elongation factor TFIIS. *Embo J* **23**(21): 4232-4242.

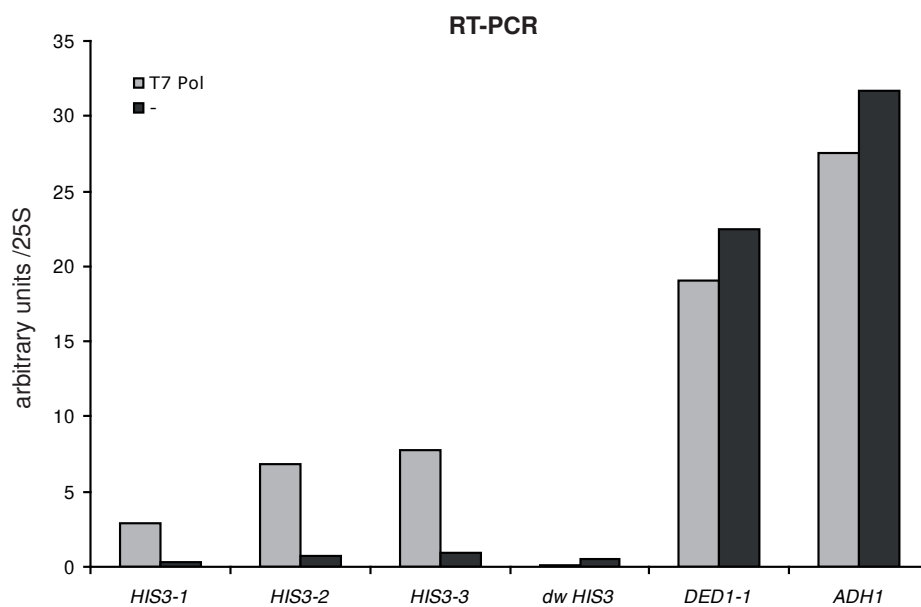
A

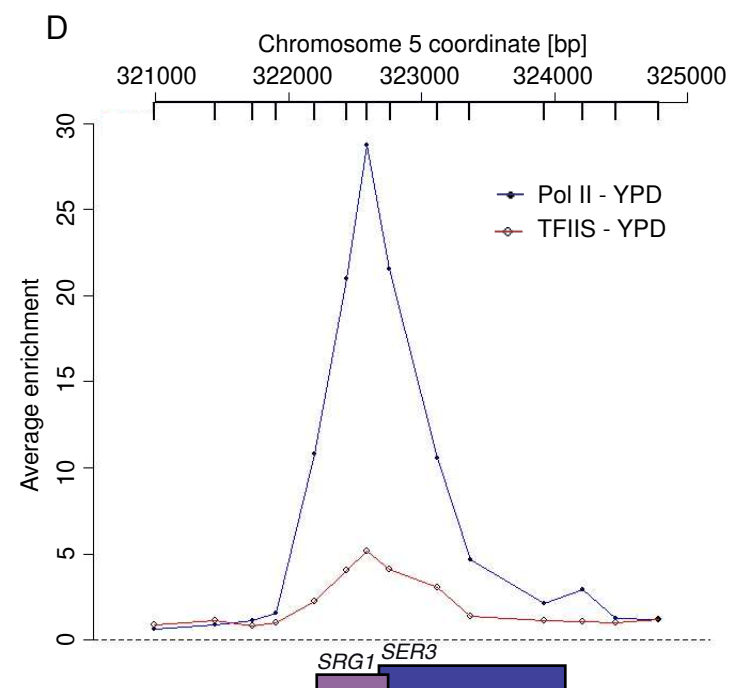
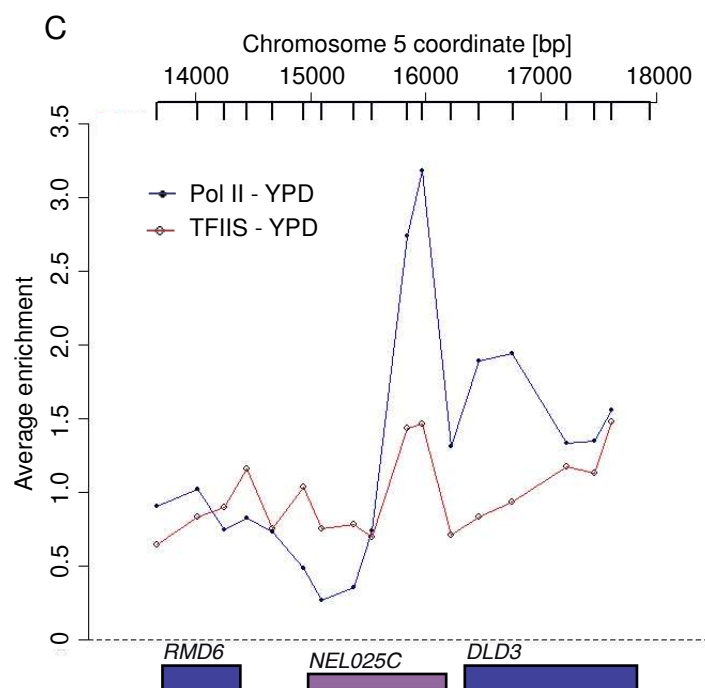
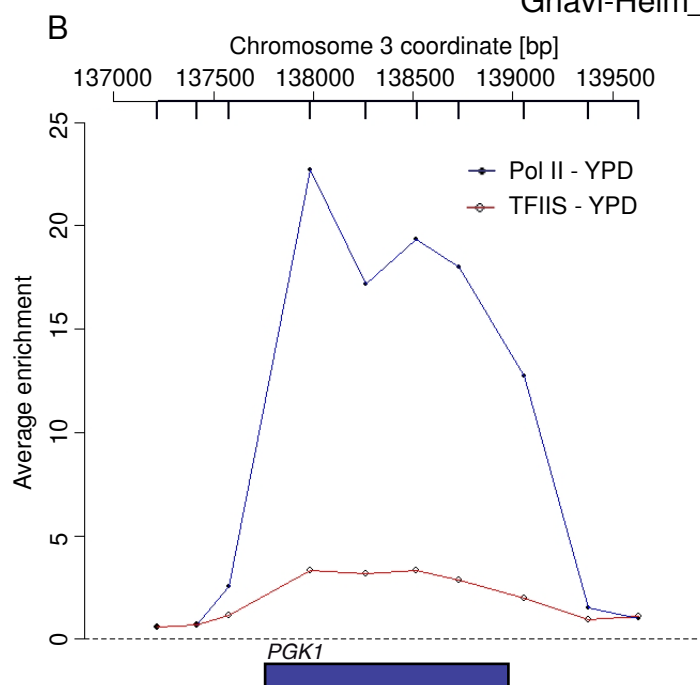
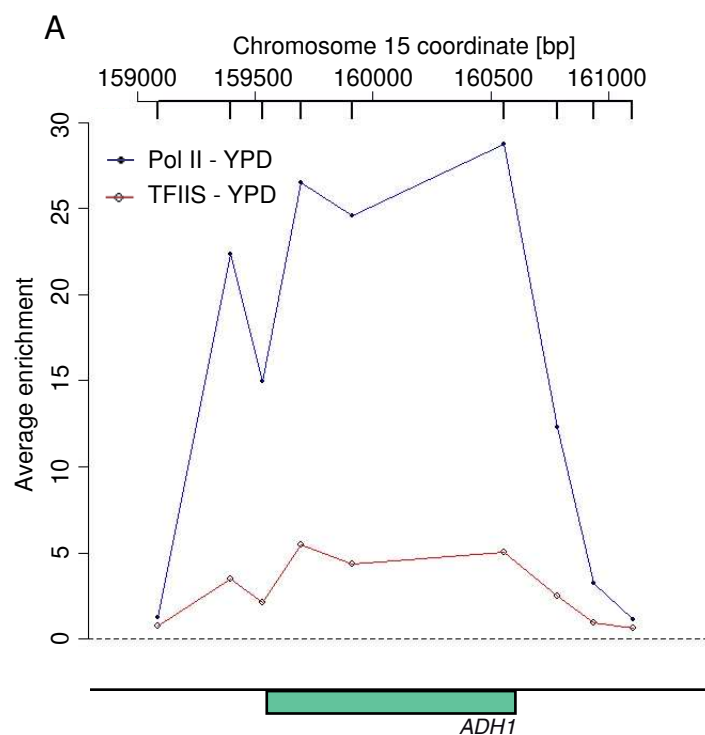


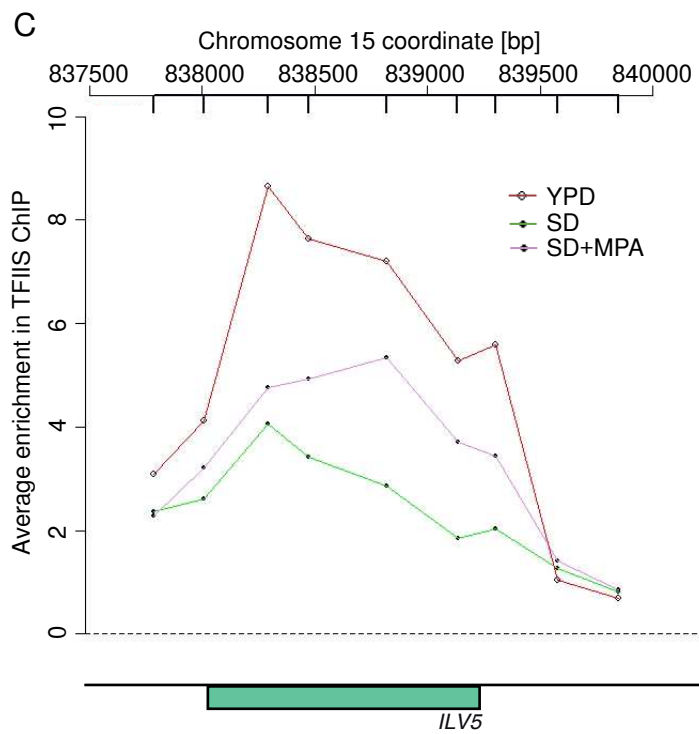
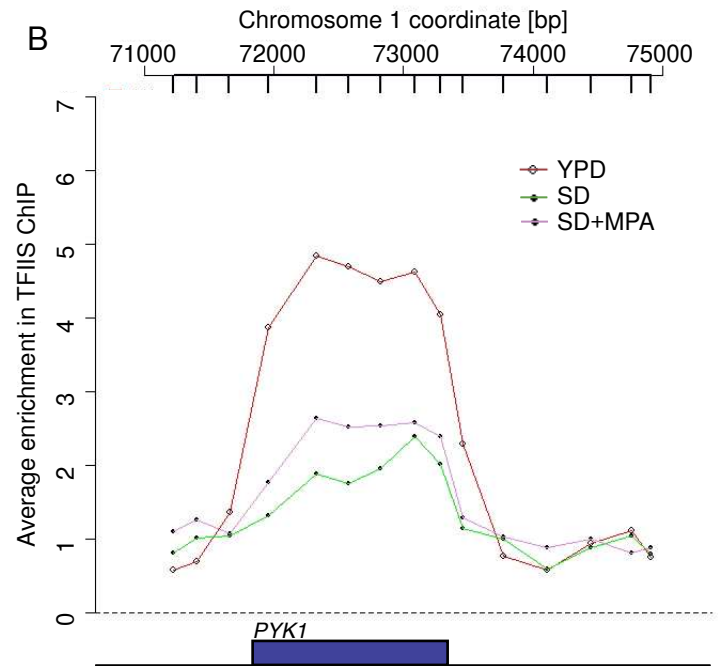
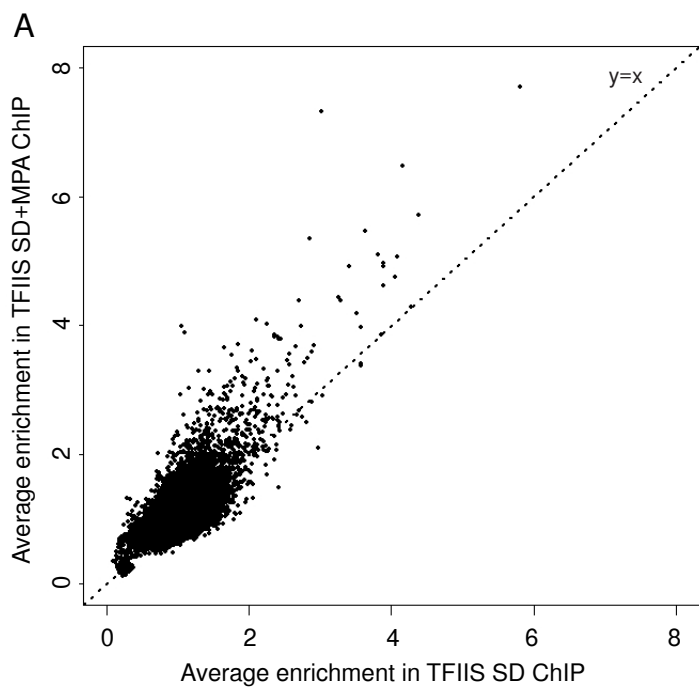
B

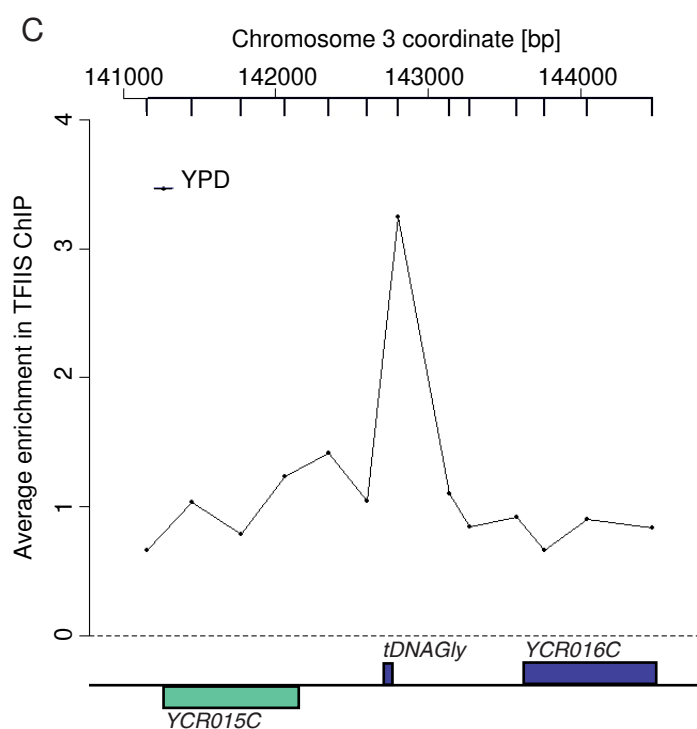
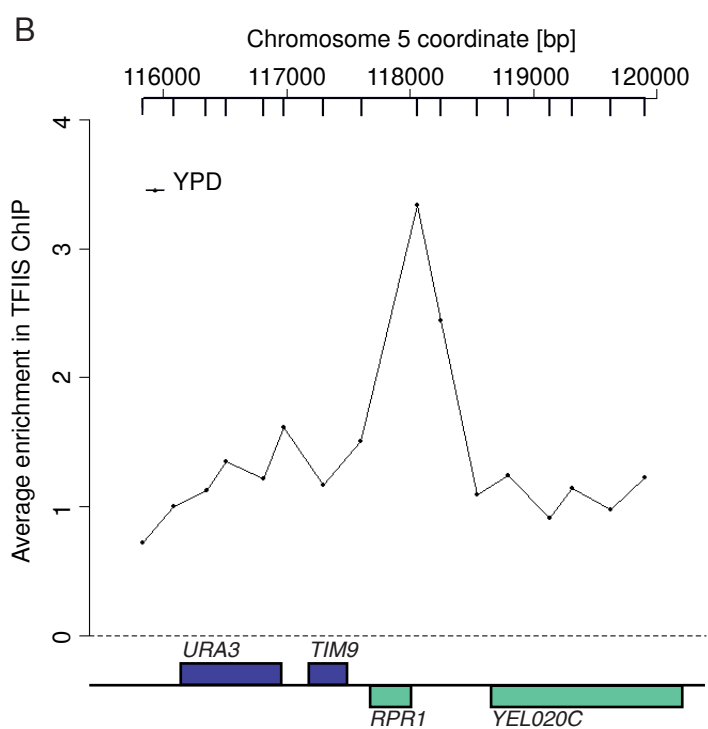
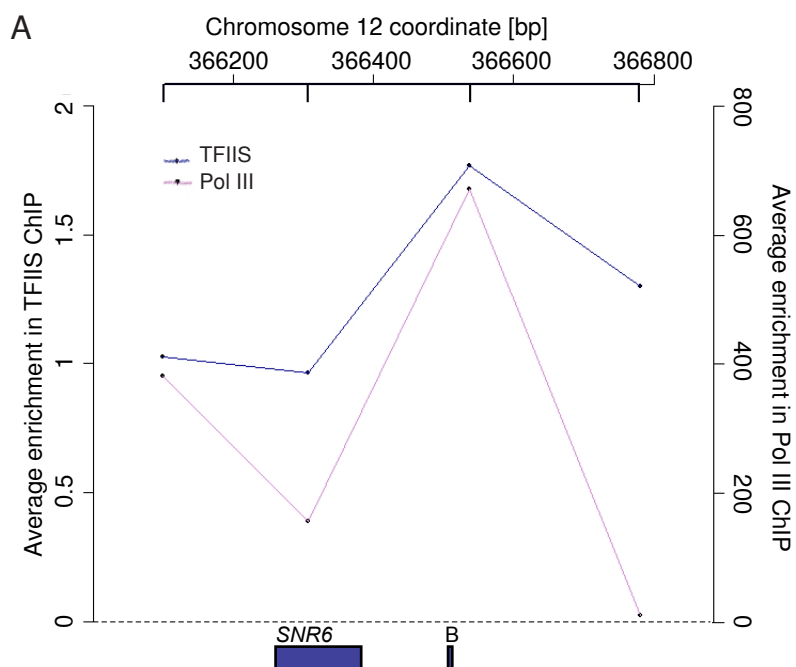


C

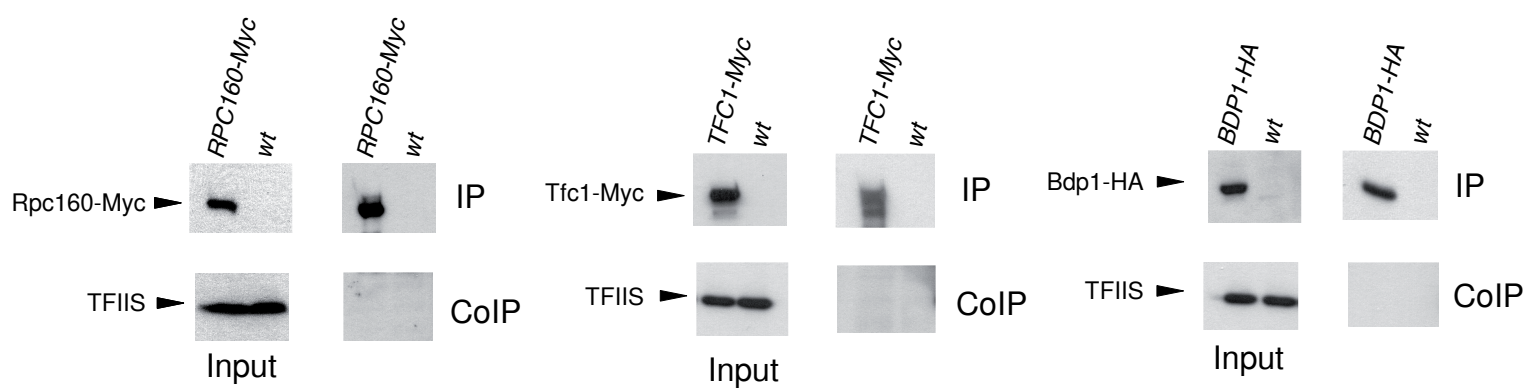




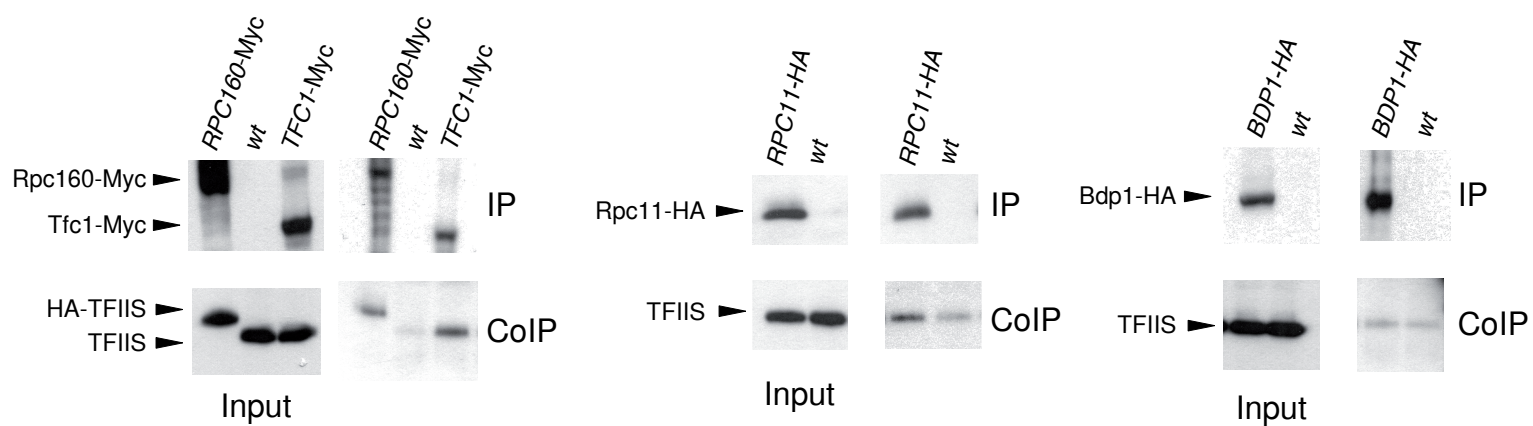


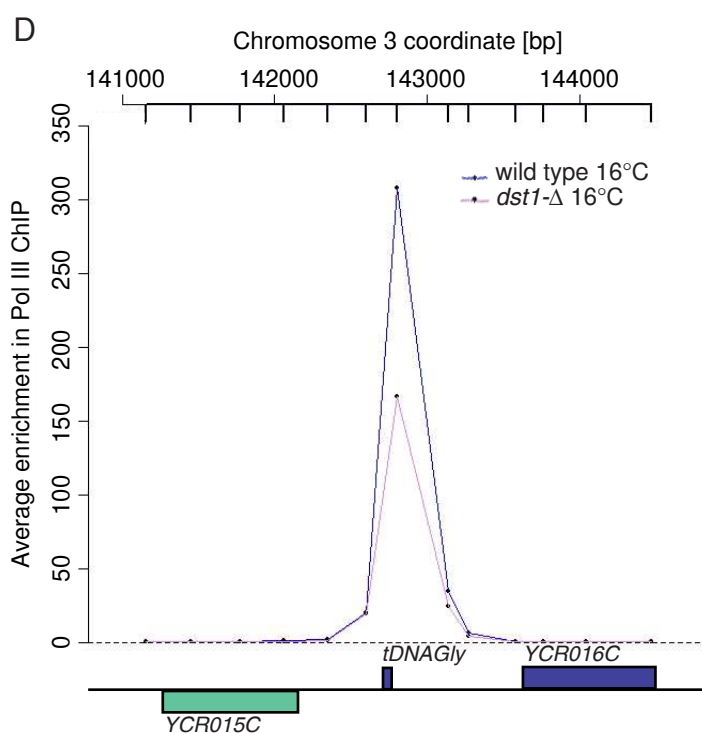
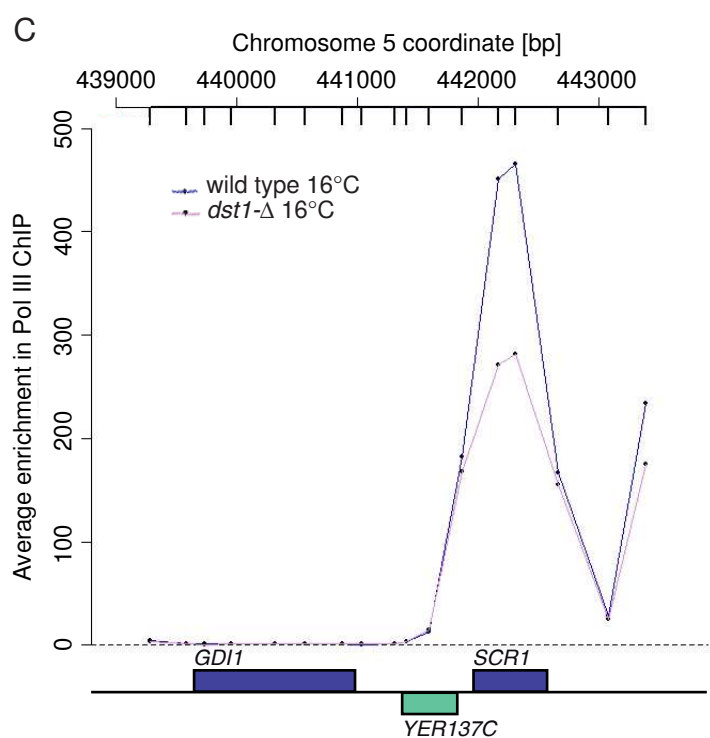
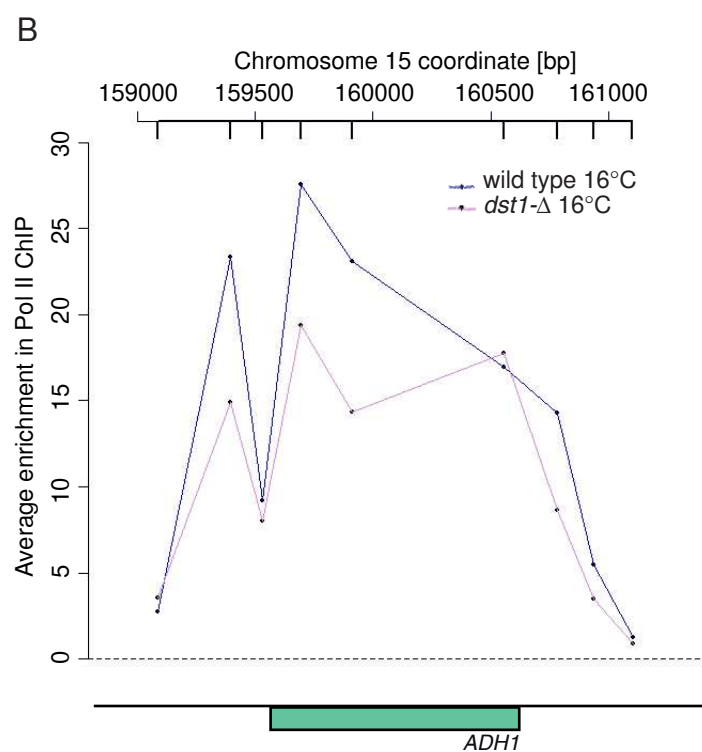
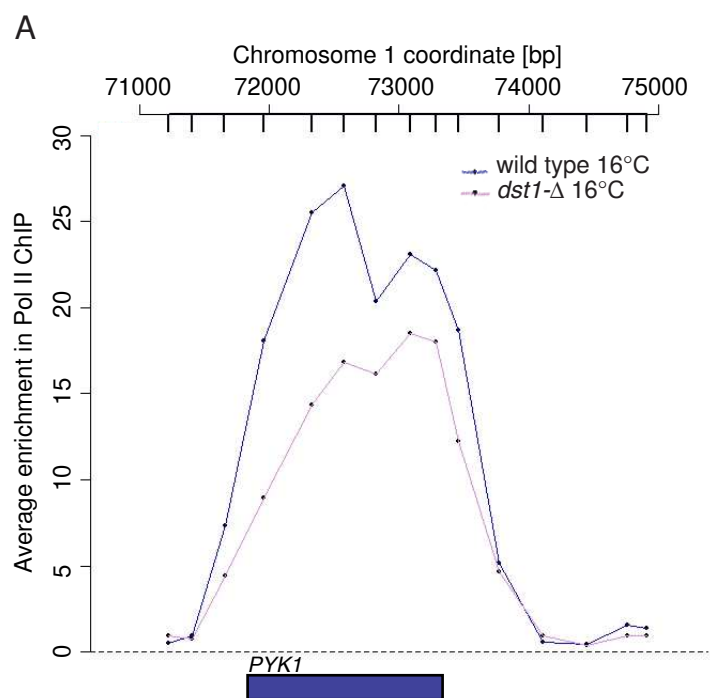


## A Before crosslink



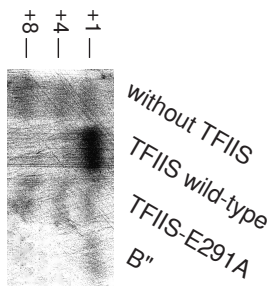
## B After crosslink



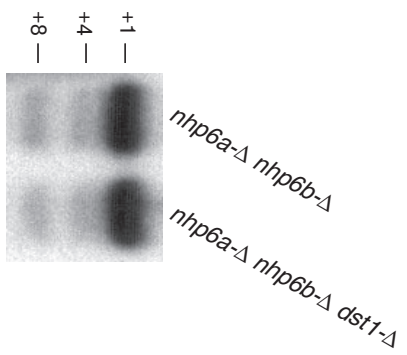


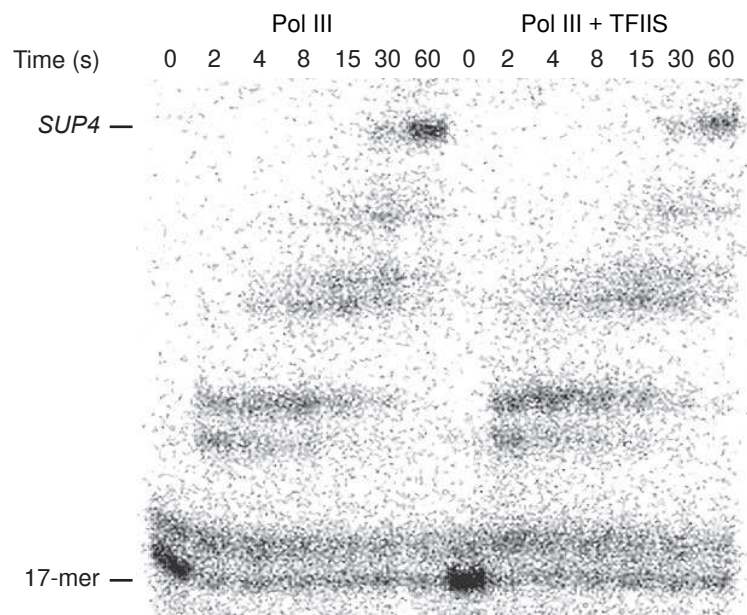


A



B



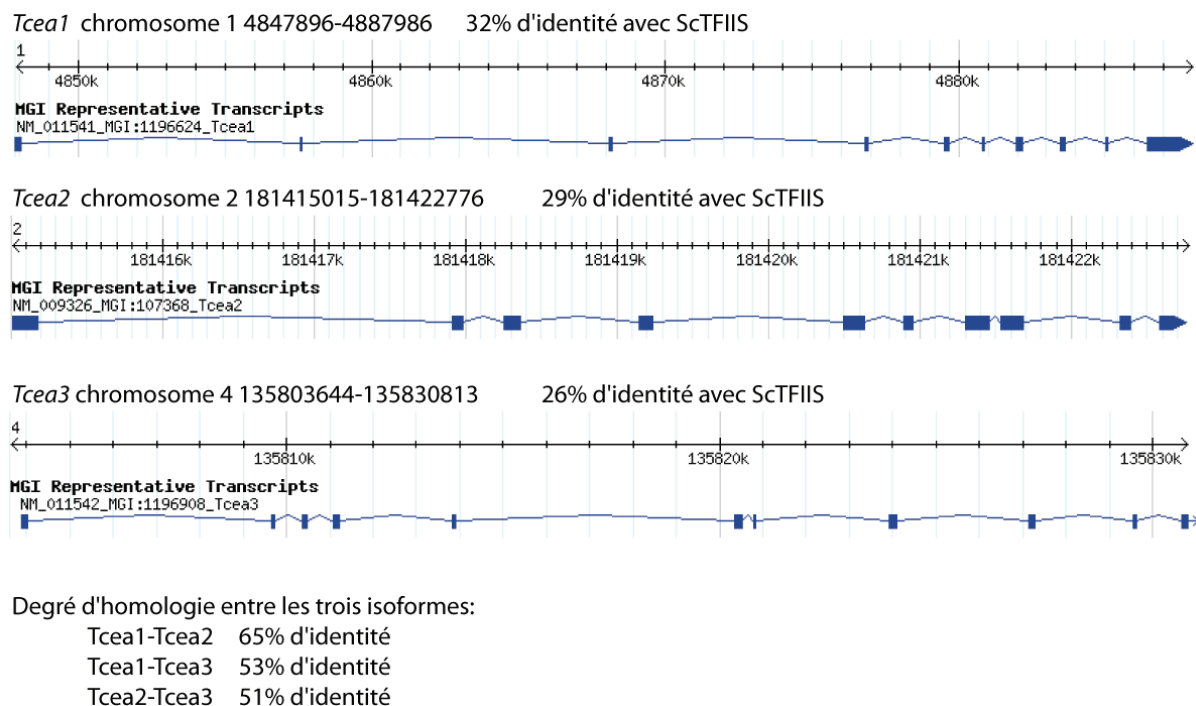




## **b) Rôle des homologues murins de TFIIS dans la régulation de la transcription chez *Mus musculus***

### Contexte du travail

A la suite de notre travail sur le facteur TFIIS chez *S. cerevisiae*, nous avons voulu étendre l'étude du rôle de ce facteur aux eucaryotes pluricellulaires, et notamment à la souris. Il existe 3 isoformes de TFIIS chez la souris appelés Tcea1, 2 et 3. Ces trois protéines sont relativement mal caractérisées. Tcea1 semble être exprimée de façon ubiquitaire, Tcea2 est spécifique des spermatocytes (Ito et al., 1996), et Tcea3 est exprimé dans le foie, les reins et le cœur (Labhart & Morgan, 1998). L'organisation des trois gènes codant ces protéines est représentée dans la Figure 25. Nous avons voulu étudier la localisation à grande échelle de ces isoformes par des expériences de ChIP-seq (immunoprécipitation de la chromatine suivie d'un séquençage à haut débit). Des anticorps de qualité n'étant pas disponible pour l'immunoprécipitation de ces protéines, une première étape a consisté à construire des lignées des cellules souches embryonnaires (ES) de souris étiquetées pour chacune de ces isoformes. Nous pourrions ainsi déterminer si le rôle de TFIIS dans la transcription par l'ARN polymérase III est conservé au cours de l'évolution. Pour ce faire, nous comparerons la localisation de TFIIS à celle de l'ARN polymérase II et III (expériences en cours dans le laboratoire de Pierre Ferrier (Centre d'Immunologie Marseille-Luminy) et le nôtre, respectivement). Nous pourrions également déterminer si les trois isoformes de TFIIS ont des rôles différents dans les cellules ES, ou si une des formes est majoritairement utilisée, les autres étant davantage tissu-spécifiques. A plus long terme, nous pourrions produire des lignées de souris à partir des cellules ES. Ces expériences pourront mieux nous éclairer sur le rôle de TFIIS *in vivo* dans la transcription par l'ARN polymérase II et III.



**Figure 25: Organisation chromosomique des gènes *Tcea1*, *Tcea2* et *Tcea3*.**

L'organisation en intron et exons de chacun des trois gènes a été obtenu d'après la base de données MGI (Mouse Genome Informatics, <http://www.informatics.jax.org/>). Le chromosome, la position, et le pourcentage d'identité avec la protéine TFIIS de *S. cerevisiae* (ScTFIIS) sont indiqués. Le pourcentage d'identité entre chacune de ces isoformes est également mentionné.

## Résultats

### - Construction des vecteurs de recombinaison

La technique du "Recombineering" (Liu et al., 2003) a récemment été importée et développée au laboratoire dans le cadre d'une collaboration avec le laboratoire de Matthieu Gérard (CEA/Saclay) pour l'étude de la localisation globale de la machinerie de transcription par l'ARN polymérase III. Cette technologie permet de générer rapidement et avec une haute fidélité les vecteurs de recombinaison homologue qui ciblent l'introduction de la séquence codant une étiquette avant le premier ou dans le dernier exon du gène codant une protéine d'intérêt.

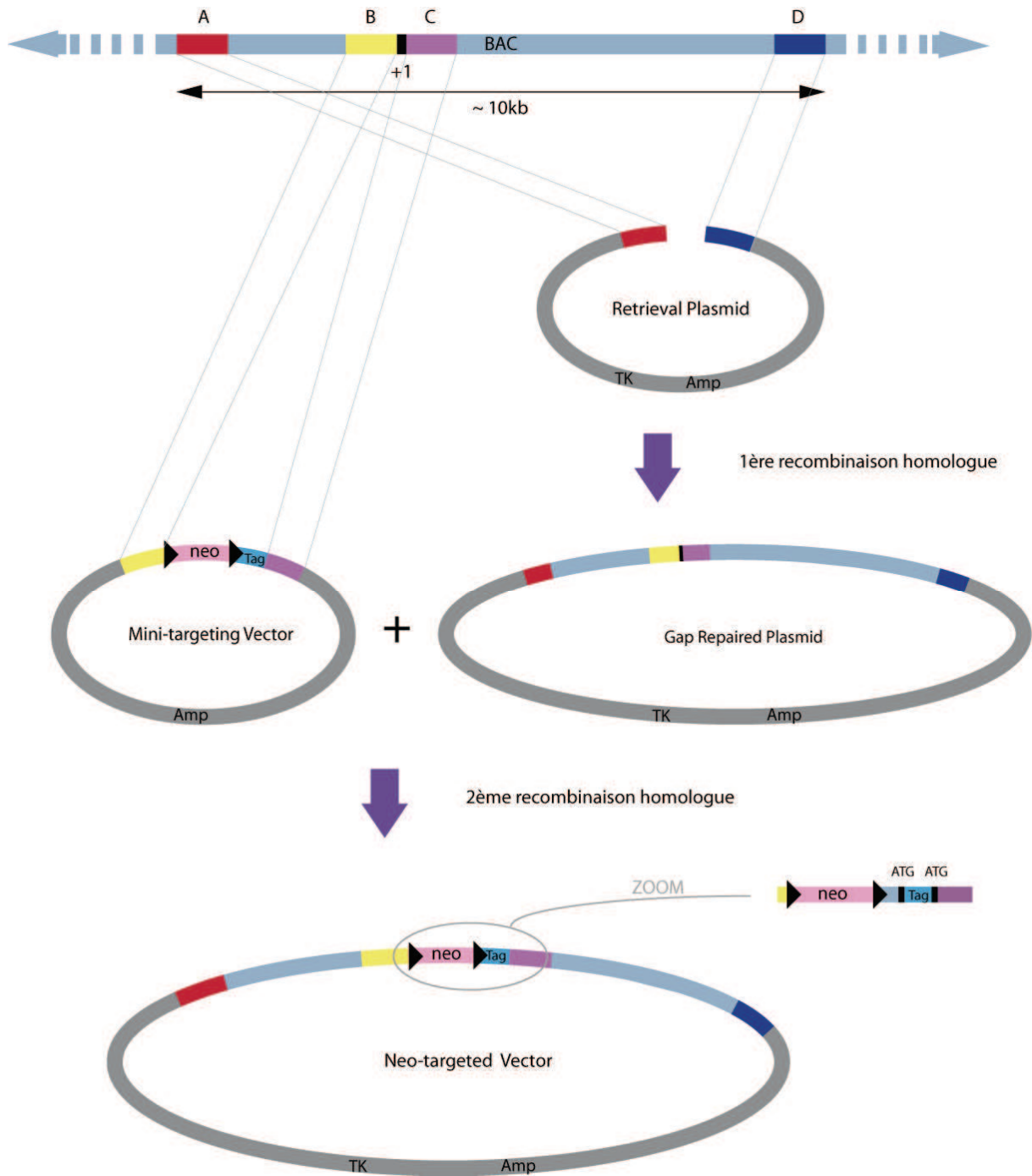
Afin d'optimiser les procédés de purification des protéines ainsi étiquetées (ou "tagguées"), nous avons choisi une étiquette de 33 acides aminés construit et utilisé avec succès par le laboratoire de V. Ogryzko, qui comprend la séquence suivante: 6His-Flag-HA.

Cette étiquette est introduite en N-terminal des isoformes de TFIIS, car l'ajout d'une étiquette en C-terminal chez *S. cerevisiae* conduit à l'expression d'une protéine non fonctionnelle. Un récapitulatif de la construction des vecteurs de recombinaison est présenté en Figure 26. Brièvement, quatre séquences de 250 à 500 paires de base appelées bras proximaux 5'/3' (de part et d'autre du +1 de traduction) et bras distaux 5'/3' (définissant une région d'environ 10 kb autour du +1 de traduction) sont choisies. Ces séquences sont ensuite amplifiées par PCR à partir d'un BAC comprenant notre région d'intérêt et clonées dans le plasmide pL253 pour les bras distaux 5'/3' ou dans un plasmide pL452 modifié comportant notre étiquette en aval de la cassette de résistance neo pour les bras proximaux 5'/3' (voir description des plasmides en Annexe). L'utilisation d'une souche d'*E. coli* (SW102) modifiée comportant des protéines du phage  $\lambda$  permet la construction du "Gap Repaired Plasmid" grâce à une première recombinaison homologue entre le BAC et le "Retrieval Plasmid". Une seconde étape de recombinaison homologue entre le "Gap Repaired Plasmid" et le "Mini-targeting Vector" permet d'obtenir un vecteur, le "Neo-targeted Vector" comportant l'étiquette en aval de la cassette neo.

Ces vecteurs de recombinaison ont pu être construits pour les trois gènes *Tcea1*, *Tcea2* et *Tcea3*. Les constructions et les oligonucléotides utilisés sont décrits dans la Figure 27 et le Tableau 6, respectivement. Il est à noter que le séquençage du vecteur de recombinaison de *Tcea3* révèle la présence de nombreuses mutations silencieuses (mutations dans les régions introniques ou modification d'une base sans changement de l'acide aminé). Ces mutations sont répertoriées par la base de données MGI (Mouse Genome Informatics, <http://www.informatics.jax.org/>) comme des zones polymorphisme avérées. Bien qu'elles ne modifient pas la séquence de notre protéine d'intérêt, ces mutations pourraient avoir un impact sur l'efficacité de la recombinaison homologue dans les cellules ES.

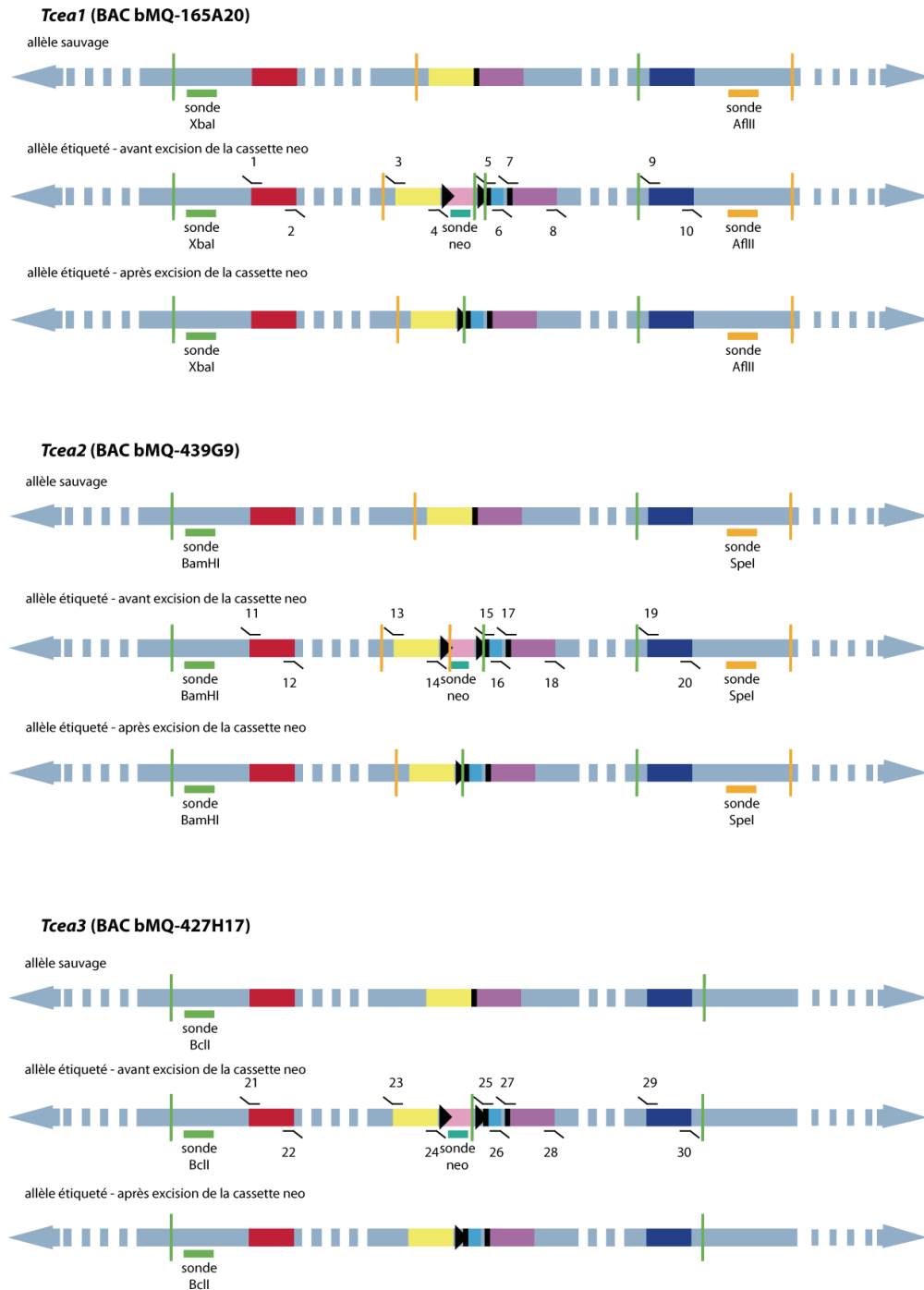
#### - Obtention de lignées de cellules ES étiquetées

Notre équipe utilise pour toutes ses expériences la lignée de cellules ES 46C, qui possède un allèle du gène *Sox1* fusionné avec une séquence GFP-ires-Puromycin phosphotransférase (Aubert et al., 2003). Comme *Sox1* est un marqueur précoce de la différenciation neuronale, la présence de cet allèle permet de sélectionner, en présence de puromycine, les cellules qui sont effectivement engagées dans la différenciation neuronale au cours des protocoles de différenciation des cellules ES. Cette lignée cellulaire présente la caractéristique de pouvoir être cultivée sans fibroblastes embryonnaires murins (MEFs) permettant ainsi de préparer la chromatine à partir d'un type cellulaire unique.



**Figure 26. Schéma de la technique d'étiquetage de protéines par "Recombineering".**

Le bras distal 5' (A) est représenté en rouge, le bras distal 3' (D) en bleu foncé, le bras proximal 5' (B) en jaune, le bras proximal 3' (C) en violet, la cassette neo en rose, l'étiquette (Tag) en bleu clair. Les codons ATG sont représentés par des rectangles noirs et les sites *loxP* par des flèches noires.



**Figure 27. Schéma de la construction de lignées cellulaires étiquetées pour Tcea1, Tcea2 et Tcea3.**

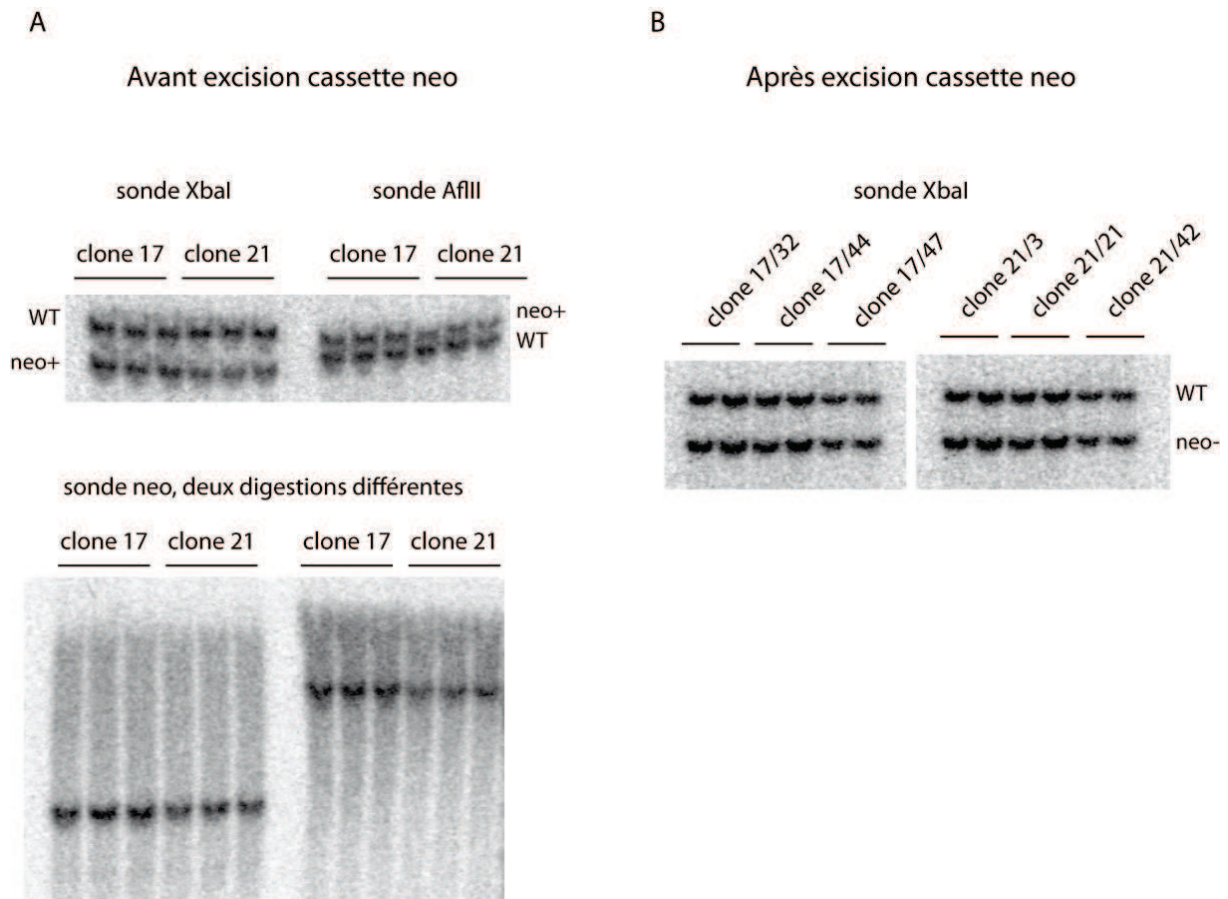
Le bras distal 5' est représenté en rouge, le bras distal 3' en bleu foncé, le bras proximal 5' en jaune, le bras proximal 3' en violet, la cassette neo en rose, l'étiquette (Tag) en bleu clair. Les codons ATG sont représentés par des rectangles noirs et les sites *loxP* par des flèches noires. Les sondes utilisées pour le génotypage sont représentées par des rectangles verts ou orange, les sites de digestion correspondants par des traits de la même couleur. Les oligonucléotides utilisés pour la construction des vecteurs de recombinaisons sont indiqués par des chiffres, la séquence correspondante est reportée dans le Tableau 6.



**Tableau 6: Oligonucléotides utilisés pour la construction des vecteurs de recombinaison.**

Oligonucléotides	Forward	Reverse
1/2	AAGGAAAAAAGCGGCCGCAAAAGG	GACTAGTCATTTGATACATGCCCCCT
3/4	AAAATTCAGCTCGAGGCTTAGCA ACGCGTCGACGTCGGCCATAGCGGC GCGGAAGGGTTTCGATGGAACTCGT A	CA GGAATTCCGGCTCCGGAAGGCCTCCC CT
5/6, 15/16, 25/26	CGGGATCCCGTCTAGATGCATCACCA TCACCACCATGG	AAGGAAAAAAGCGGCCGCAAAAGG AAAAGCTTGCGTAGTCGGGCACGT C
7/8	CCCAAGCTTATGGAGGACGAGGTGG TTCGCATTGCCAAG	TTTTCTTTTTCGCGCCGTCAGTGTGGA AGCAAAGGGATGGAAAGTGGTA
9/10	GACTAGTCGGCACTGAGTATGCATGT GG	CGGGATCCCGATGGCAGAGGACGAC TTTC
11/12	AAGGAAAAAAGCGGCCGCAAAAGG AAATTCTTGGCTCCAAATCTTGC	GACTAGTCGGTGGGACTCTTTGACTC CA
13/14	ACGCGTCGACGTCGGCCATAGCGGC CGCGGAAGCTGCTCAGATGGGTTGA AT	GGAATTCCCACTACCTCGGTGGCGGC GG
17/18	CCCAAGCTTATGGGCAAGGAGGAAG AGAT	TTTTCTTTTTCGCGCCGCTTTTTTCCT TGATGCTCCTCAAAGGCTACG
19/20	GACTAGTCCAACCCAGAGCCTGTTTC TC	CGGGATCCCTCCCTTGATTCTGAT TTG
21/22	AAGGAAAAAAGCGGCCGCAAAAGG AAACCACCATCGCTTAGTCCAAT	GACTAGTCCCTTGCCTCACCAAATGT CT
23/24	CGGGGTACCCCGGACGTCAAATCT GCACCAA	ACGCGTCGACGTTGGTCCACAACGC ACGGC
27/28	CCCAAGCTTATGGGCCTGGAAGAGG AGCT	TTTTCTTTTTCGCGCCGCTTTTTTCCT TCGAACCCCACTCCTAA
29/30	GACTAGTCGTATAGCCCTGGCTGTCC TG	CGGGATCCCTAGCCACAAACCCAT GGTC

Nous avons transformé par électroporation les "Neo-targeted Vector" linéarisés obtenus pour les trois constructions dans des cellules ES 46C. Les clones positifs sont sélectionnés de par leur résistance au G418 (apportée par la cassette neo) et au gancyclovir (contre sélection de l'intégration de la cassette TK codant la thymidine kinase, une enzyme capable de modifier le gancyclovir en un produit toxique). Une centaine de clones par construction sont généralement repiqués, amplifiés puis génotypés par Southern blot pour vérifier l'intégration de la séquence codant l'étiquette au locus (Figure 28A). Les sondes et digestions utilisées pour le génotype ainsi que la taille des bandes attendues sont indiquées sur les Figure 27 et 29, respectivement. Dans le cas de l'étiquetage de l'isoforme Tcea3, aucun clone positif n'a pu être obtenu malgré plusieurs tentatives d'électroporation avec différentes concentrations de "Neo-targeted Vector" issu de plusieurs préparations de plasmide différentes. Dans les cas de l'étiquetage de l'isoforme Tcea2, l'expression de la protéine s'avère trop faible dans les cellules ES pour effectuer des expériences d'immunoprécipitation de la chromatine. Cette construction pourra cependant être utilisée plus tard pour l'obtention de lignées de souris afin d'étudier le rôle de cette protéine de façon tissu-spécifique. Deux clones



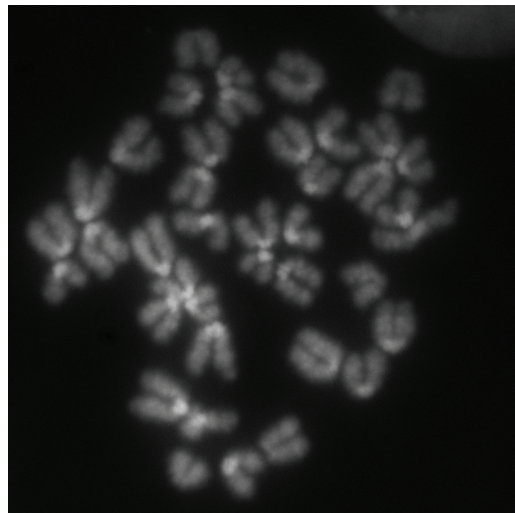
**Figure 28. Génotypage des clones de Tcea1 étiquetés par Southern blot.**

**A.** Vérification de l'étiquetage des clones 17 et 21 grâce à la sonde XbaI (panneau du haut, gauche), à la sonde AflIII (panneau du haut, droite), ou grâce à la sonde neo en utilisant deux digestions différentes (panneau du bas). L'allèle sauvage (WT) et ayant intégré la cassette neo (neo+) sont indiqués. **B.** Vérification de l'excision de la cassette neo des clones 17/32, 17/44, 17/47, 21/3, 21/21 et 21/42 grâce à la sonde XbaI. L'allèle sauvage (WT) et ayant intégré l'étiquette après excision de la cassette neo (neo-) sont indiqués. Les tailles attendues sont indiquées dans la Figure 29.

**Génotypage: tailles attendues**

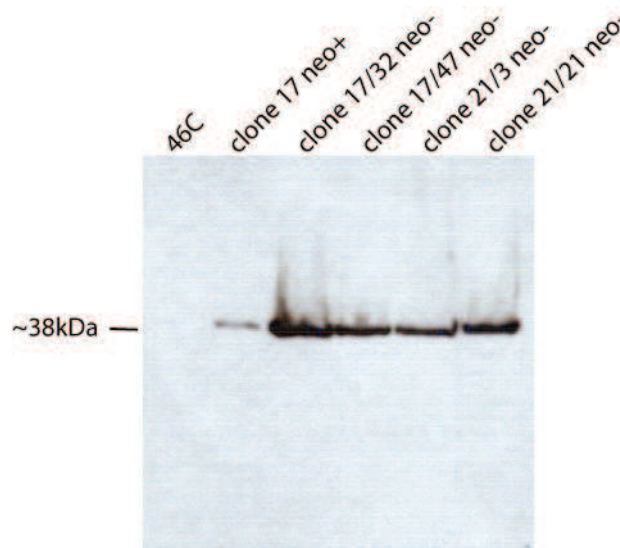
		allèle sauvage	allèle étiqueté neo+	allèle étiqueté neo-
<i>Tcea1</i>	digestion XbaI	14145 pb	9723 pb	8288 pb
	digestion AflIII	10375 pb	12386 pb	10592 pb
<i>Tcea2</i>	digestion BamHI	9514 pb	8319 pb	6522 pb
	digestion SpeI	4966 pb	6210 pb	5184 pb
<i>Tcea3</i>	digestion BclI	16662 pb	10085 pb	16901 pb

**Figure 29. Tailles des bandes attendues lors du génotypage des lignées cellulaires étiquetées pour Tcea1, Tcea2 ou Tcea3.**



**Figure 30. Observation du caryotype d'un clone positif Tcea1 étiqueté.**

Les chromosomes sont colorés au DAPI. Les figures mitotiques sont observées au microscope à fluorescence.



**Figure 31. Expression des clones positifs Tcea1 étiquetés.**

L'expression de la protéine Tcea1 étiquetée (poids moléculaire attendu d'environ 38kDa) est vérifiée par Western blot à partir d'extraits cellulaires totaux. Quatre clones positifs après excision de la cassette neo (clone 17/32 neo-, clone 17/47 neo-, clone 21/3 neo-, et clone 21/21 neo-), sont analysés par comparaison à un témoin négatif (46C, cellules ES non étiquetées), et positif (clone 17 neo+, clone positif avant excision de la cassette neo). L'intensité du signal n'est pas directement proportionnelle à l'expression de la protéine.

positifs (appelés clone 17 et clone 21) étiquetés pour l'isoforme Tcea1 ont été sélectionnés. L'expression de la protéine a été vérifiée par Western blot, puis la cassette neo a été excisée par électroporation concomitante d'un plasmide exprimant la Cre recombinase et un plasmide de résistance à la puromycine. Une cinquantaine de clones positifs, sélectionnés de par leur résistance à la puromycine, sont repiqués, amplifiés, puis testés par génotypage pour vérifier l'excision de la cassette neo (Figure 28B). Quatre clones positifs (appelés 17/32, 17/47, 21/3 et 21/21) ont été sélectionnés. L'intégrité de leur matériel chromosomique a été vérifiée par caryotypage (Figure 30), et l'expression de la protéine Tcea1 étiquetée contrôlée par Western blot grâce à des anticorps anti-HA (Figure 31). Malgré plusieurs tentatives, l'épitope Flag ne semble quand à lui pas être reconnu par les anticorps anti-Flag utilisés. Par la suite, nous utiliserons les clones 17/32 et 21/3 pour les expériences d'immunoprécipitation de la chromatine.

- Immunoprécipitation de chromatine (ChIP)

Le savoir faire pour cette technique est déjà acquis au sein du laboratoire (Harismendy et al., 2003; Schawalder et al., 2004, Ghavi-Helm et al., 2008, Esnault et al., 2008). De plus, de récentes expériences d'immunoprécipitation de la machinerie de transcription par l'ARN polymérase III réalisées au laboratoire nous ont permis d'adapter la technique pour les cellules ES. Des résultats préliminaires ont été obtenus pour la préparation de la chromatine de la lignée 46C étiquetée pour Tcea1. Cependant, nous devons encore améliorer les conditions d'immunoprécipitation de la chromatine avant de pouvoir envoyer nos échantillons au séquençage sur la plateforme Solexa.



## ***II/ Rôle du Médiateur dans le recrutement du complexe TFIID lors de la formation du PIC in vivo***

### Contexte du travail

Une première version de ce travail a été présentée par Cyril Esnault lors de sa thèse de doctorat (2007), sous forme d'un manuscrit soumis à publication. La version publiée de ce travail (Esnault et al., 2008) s'appuie sur la réalisation et l'analyse des données de localisation globale du Médiateur, de TFIID, de l'ARN polymérase II totale et phosphorylée sur la sérine 5 dans une souche sauvage et dans le mutant *med11-T47A*, dans lesquelles j'ai été plus particulièrement impliquée.

L'ARN polymérase II est incapable d'initier seule la transcription. Elle nécessite la présence des facteurs généraux. L'initiation de la transcription commence par la mise en place du complexe de préinitiation (PIC), qui apparaît comme une étape séquentielle *in vitro*. Cependant, la présence de complexes coactivateurs pourrait influencer l'assemblage du PIC *in vivo*.

Le Médiateur qui a été initialement purifié chez *S. cerevisiae*, permet de moduler le niveau de transcription en réponse aux signaux de régulation positif ou négatif. *In vitro*, il est nécessaire à l'activation de la transcription en réponse aux activateurs spécifiques, il augmente le niveau de la transcription basale et stimule l'activité kinase de TFIID (Kim et al., 1994). *In vivo*, le Médiateur est requis pour la transcription de la quasi totalité des gènes de *S. cerevisiae* (Holstege et al., 1998). Les analyses temporelles de mise en place du PIC montrent qu'il lie le promoteur des gènes avant les facteurs généraux (Bryant and Ptashne, 2003; Park et al., 2001). De plus, l'altération de son recrutement ou une mutation affectant son intégrité empêche la formation du PIC, suggérant ainsi un rôle essentiel du Médiateur dans la mise en place du complexe de préinitiation (Bhaumik et al., 2004; Takagi et al., 2006). Enfin, une étude à grande échelle d'interaction génétique montre des liens du module de Tête du Médiateur avec l'ARN polymérase II, TFIIF et TFIID (Collins et al., 2007).

Afin de découvrir comment le Médiateur peut augmenter le niveau de la transcription basale, stimuler la transcription et la phosphorylation du CTD par TFIID, et comment il peut influencer la mise en place du PIC, nous avons étudié le rôle d'une sous-unité essentielle et conservée du module de Tête du Médiateur: Med11. Une étude précédente réalisée au

laboratoire en collaboration avec l'équipe de F. Holstege (Utrecht) a permis de dresser une carte des interactions au sein du Médiateur (Guglielmi et al., 2004), mais aussi des interactions des sous-unités du Médiateur avec d'autres protéines nucléaires. Nous avons ainsi pu identifier par double-hybride une interaction entre Med11 et un domaine conservé de Rad3, une hélicase du cœur du complexe TFIIF. Nous avons ensuite confirmé cette interaction par co-immunoprécipitation (CoIP) à partir d'extraits cellulaires.

Pour étudier le rôle de cette interaction et la fonction de Med11, nous avons réalisé une mutagenèse aléatoire de la protéine suivie d'un crible de thermosensibilité. Elle nous a permis d'obtenir six mutants ponctuels, répartis sur l'ensemble de la protéine et deux mutants de troncature. Nous avons alors caractérisé ces mutants par double-hybride entre Med11 et ses partenaires: Med17, Med22 et Rad3. Nous avons ainsi identifié trois mutants particulièrement intéressants pour la compréhension des mécanismes mis en jeu par le Médiateur: le mutant *med11-T47A* qui est spécifiquement affecté pour l'interaction avec Rad3 (TFIIF), le mutant *med11-G108S* qui est spécifiquement affecté pour l'interaction avec Med17 (Médiateur) et le mutant *med11-L82P* qui est affecté pour toutes les interactions testées (Rad3, Med17 et Med22). Après avoir confirmé le défaut d'interaction entre Med11-T47A et Rad3 par CoIP, nous avons vérifié dans les trois mutants (*med11-T47A*, *-L82P* et *-G108S*), l'intégrité du Médiateur et la capacité du complexe à interagir avec l'ARN polymérase II. Nous avons observé que seul le mutant présentant un défaut pour toutes les interactions connues de Med11 (*med11-L82P*), affecte l'intégrité du module. Ceci est en accord avec le rôle central de Med11 dans la structure du module Tête (Takagi et al., 2006). Par contre les trois mutants sont toujours capables d'interagir avec l'ARN polymérase II. Néanmoins, une analyse des effets de ces mutations sur la transcription par RT-PCR ("Reverse Transcription Polymerase Chain Reaction") montre qu'à 37°C, ces trois mutants sont affectés dans la quantité des ARN messagers testés.

Ces trois mutants nous ont ensuite permis d'étudier le rôle du Médiateur dans la mise en place du PIC. Après nous être assurés qu'ils n'affectent pas le niveau d'expression de protéines clefs de la transcription, nous avons étudié par ChIP l'assemblage du PIC. La protéine TBP et le Médiateur se fixent normalement sur le promoteur des gènes étudiés. Par contre, les trois mutations réduisent l'occupation de l'ARN polymérase II, TFIIE et TFIIF. De façon surprenante, contrairement à la séquence d'assemblage établie *in vitro*, le mutant *med11-G107S* qui est spécifiquement affecté dans l'interaction entre Med11 et Med17, présente un défaut d'occupation de l'ARN polymérase II sans affecter ceux de TFIIE et TFIIF. Ceci suggère un recrutement indépendant de ces trois facteurs. Le mutant *med11-*

*T47A* qui est affecté dans l'interaction avec Rad3, réduit spécifiquement le niveau d'occupation du module TFI<sub>IK</sub>, mais pas du cœur de TFI<sub>IIH</sub>, suggérant que l'interaction entre le Médiateur et Rad3 est importante pour la stabilisation et l'association des sous-complexes de TFI<sub>IIH</sub>. Enfin, la mutation *med11-L82P* qui affecte toutes les interactions testées, réduit l'occupation de l'ARN polymérase II, TFI<sub>IE</sub> et de l'ensemble du complexe TFI<sub>IIH</sub> sur la chromatine. Ces trois mutants indiquent que le Médiateur influence la mise en place du PIC en favorisant indépendamment l'occupation de l'ARN polymérase II et de TFI<sub>IIH</sub>/TFI<sub>IE</sub> et en stabilisant l'association entre TFI<sub>IK</sub> et le cœur de TFI<sub>IIH</sub>.

L'effet de la mutation *med11-T47A* sur l'occupation du module TFI<sub>IK</sub> semble promoteur-spécifique. En effet, sur cinq gènes testés par ChIP, seuls deux gènes, *ADH1* et *PYK1* semblent être affectés. Nous avons donc analysé l'effet de la mutation *med11-T47A* sur l'occupation de la sous-unité Kin28 du module TFI<sub>IK</sub> à l'échelle du génome afin d'avoir une idée plus précise de la proportion de gènes affectés. Des expériences de ChIP-on-chip ont révélé que seuls 665 gènes présentent une diminution de deux fois ou plus de la présence de Kin28 dans la souche *med11-T47A* par rapport à une souche sauvage. Nous avons vérifié que ces gènes étaient spécifiquement affectés pour leur occupation par TFI<sub>IK</sub>, mais pas pour celle de la sous-unité Rad3 du cœur de TFI<sub>IIH</sub> ou de la sous-unité Med5 du Médiateur.

TFI<sub>IIH</sub> phosphoryle le CTD de l'ARN polymérase II par l'intermédiaire de son module TFI<sub>IK</sub>. Cette étape est essentielle dans la transition entre l'initiation et l'élongation de la transcription. Nous avons donc analysé l'effet, à l'échelle du génome, de la mutation *med11-T47A* sur la phosphorylation du CTD, *in vivo*. Nous avons ainsi pu remarquer, en accord avec la baisse du niveau d'occupation du module TFI<sub>IK</sub>, que le niveau de phosphorylation du CTD sur la sérine 5 diminuait de façon globale dans le mutant *med11-T47A*, sans que le niveau d'ARN polymérase II totale soit significativement affecté. Cet effet a été confirmé par des expériences de ChIP sur les gènes *HST2*, *ADH1* et *PYK1*. Une des propriétés du Médiateur est de stimuler l'activité kinase de TFI<sub>IK</sub> *in vitro* indépendamment de Rad3 (Guidi et al., 2004; Nair et al., 2005). Nous nous sommes alors assuré qu'*in vitro* le mutant *med11-T47A* n'était pas affecté dans cette activité et que le défaut du mutant n'affectait que le recrutement de TFI<sub>IK</sub>.

Enfin, l'observation selon laquelle la mutation *med11-G107S* réduit l'occupation de l'ARN polymérase II sans affecter celle du Médiateur ou de TFI<sub>IIH</sub> suggère que le Médiateur pourrait être un déterminant important pour le recrutement de TFI<sub>IIH</sub> sur les promoteurs. Afin de confirmer cette hypothèse, nous avons réalisé des expériences de localisation à grande échelle de l'ARN polymérase II, du cœur de TFI<sub>IIH</sub> (Rad3), du module TFI<sub>IK</sub> (Kin28) et du



Médiateur (Med5) dans une souche sauvage, et recherché des promoteurs liés par TFIID et le Médiateur mais pas par l'ARN polymérase II. Parmi les régions identifiées suivant ce profil, nous avons pu noter la présence des 16 centromères qui, à priori, ne sont pas transcrits chez *S. cerevisiae*.

L'ensemble de ces résultats indique que le Médiateur influence l'assemblage du PIC en favorisant l'occupation de l'ARN polymérase II indépendamment de TFIID et TFIIE, et en stabilisant le complexe TFIID. Il permettrait ainsi de stimuler la transcription et la phosphorylation du CTD *in vivo*, probablement en stabilisant TFIID sur les promoteurs. Ces résultats nous ont permis de proposer un modèle de la formation de PIC *in vivo*, différant fortement du modèle basé sur des expériences *in vitro* en absence du Médiateur, dans lequel plusieurs voies d'assemblage pourraient coexister.

#### Article

# Mediator-Dependent Recruitment of TFIID Modules in Preinitiation Complex

Cyril Esnault,<sup>1,4</sup> Yad Ghavi-Helm,<sup>1</sup> Sylvain Brun,<sup>1,3</sup> Julie Soutourina,<sup>1</sup> Nynke Van Berkum,<sup>2</sup> Claire Boschiero,<sup>1</sup> Frank Holstege,<sup>2</sup> and Michel Werner<sup>1,\*</sup>

<sup>1</sup>CEA, iBiTec-S, F-91191 Gif-sur-Yvette cedex, France

<sup>2</sup>Department of Physiological Chemistry, University Medical Center Utrecht, Universiteitsweg 100, 3584 CG Utrecht, The Netherlands

<sup>3</sup>Present address: Institut de Génétique et Microbiologie, UMR 8621, Université Paris Sud, F-91405 Orsay cedex, France

<sup>4</sup>Present address: Transcription Laboratory, Cancer Research UK London Research Institute, Lincoln's Inn Fields Laboratories, 44 Lincoln's Inn Fields, London WC2A 3PX, UK

\*Correspondence: [michel.werner@cea.fr](mailto:michel.werner@cea.fr)

DOI 10.1016/j.molcel.2008.06.021

## SUMMARY

**In vitro**, without Mediator, the association of general transcription factors (GTF) and RNA polymerase II (Pol II) in preinitiation complexes (PIC) occurs in an orderly fashion. In this work, we explore the *in vivo* function of Mediator in GTF recruitment to PIC. A direct interaction between Med11 Mediator head subunit and Rad3 TFIID subunit was identified. We explored the significance of this interaction and those of Med11 with head module subunits Med17 and Med22 and found that impairing these interactions could differentially affect the recruitment of TFIID, TFIIE, and Pol II in the PIC. A *med11* mutation that altered promoter occupancy by the TFIID kinase module of TFIID genome-wide also reduced Pol II CTD serine 5 phosphorylation. We conclude that the Mediator head module plays a critical role in TFIID and TFIIE recruitment to the PIC. We identify steps in PIC formation that suggest a branched assembly pathway.

## INTRODUCTION

In eukaryotes, RNA polymerase II (Pol II) is responsible for the transcription of coding genes and of a large number of noncoding RNAs. The first step in transcription activation is the recognition of DNA motifs by activators that trigger the recruitment of coactivators, GTFs, and Pol II to form a preinitiation complex (PIC).

The pathway to the formation of the PIC was initially investigated *in vitro* using a reconstituted system with template DNA, purified Pol II, and GTFs (Buratowski et al., 1989; Ranish and Hahn, 1996). It was shown that a system consisting of TFIID, -B, -D, -E, -F and -H GTFs, in addition to Pol II and promoter DNA, was sufficient to drive transcription at a basal level. The order of assembly of GTFs was determined with TFIID-binding promoter DNA first, then followed by TFIID, -B, and Pol II associated with TFIIF. Finally, TFIIE and -H completed the PIC, allowing the transition from transcription initiation to elongation after

phosphorylation of the carboxy-terminal domain (CTD) of the Pol II largest subunit and by DNA strand separation by TFIID (Akoulitchiev et al., 1995; Holstege et al., 1996). In total, five subcomplexes assembled in a unique order were identified in the pathway leading to PIC formation.

In addition to its role in basal transcription, TFIID is implicated in transactivation, the cell cycle, and nucleotide excision repair (NER) (Fisher, 2005; Zurita and Merino, 2003). It is composed of ten subunits highly conserved from yeast to mammals (Giglia-Mari et al., 2004; Ranish et al., 2004). Low resolution models of the TFIID structure have been obtained for the yeast (Chang and Kornberg, 2000) and human complexes (Schultz et al., 2000). They show an overall conservation of shape with a ring topped by a protrusion. TFIID can be separated into two subcomplexes, a core TFIID (cTFIID) and TFIID/CAK (in yeast and human, respectively) containing the Kin28/Cdk7 kinase that is responsible for CTD phosphorylation (Feaver et al., 1994; Keogh et al., 2002). Core TFIID contains two helicases of opposite polarities, Rad3/XPD and Ssl2/XPB, that are both implicated in human autosomal recessive disorders such as xeroderma pigmentosum (XP), Cockayne syndrome (CS), and trichothiodystrophy (TTD). These syndromes are due to defects in TFIID, which might affect not only NER, but also transcription (Egly, 2001; Zurita and Merino, 2003). Ssl2/XPB and Rad3/XPD play crucial roles in NER and in transcription initiation (Egly, 2001; Guzder et al., 1994a; Guzder et al., 1994b). In addition, the XPD subunit bridges the core and CAK modules of TFIID (Drapkin et al., 1996; Reardon et al., 1996) and stimulates transcription by anchoring the CAK module (Tirode et al., 1999).

Even though DNA, Pol II, and GTFs are sufficient to drive basal transcription *in vitro*, they are unable to respond to activators. This observation supports the notion that a coactivator is needed for the response to DNA-binding activators. This has prompted the purification of Mediator from yeast (Flanagan et al., 1991; Kelleher et al., 1990; Kim et al., 1994). Independent genetic evidence has led to the finding that Pol II can be found associated with Mediator, suggesting that it activates transcription by recruiting the enzyme to activated promoters (Koleske and Young, 1994). Importantly, a mutation in the Med17 head subunit of Mediator shuts down the expression of nearly all yeast genes transcribed by Pol II, indicating that it is generally required for class II transcription (Holstege et al., 1998). Subsequently, Mediator has

been isolated from different animal models based on its association with transcription activators (Conaway et al., 2005). In vitro studies of the Mediator function indicated that it stimulates basal transcription and TFIH kinase activity on serine 5 of CTD. The latter is a key event for promoter clearance and for the disruption of Mediator and Pol II interaction (Kim et al., 1994; Max et al., 2007). In addition, Mediator and TFIID were found to stabilize each other in the PIC and stimulate TFIIB assembly (Baek et al., 2006; Johnson et al., 2002).

In *Saccharomyces cerevisiae*, Mediator is composed of 25 subunits (Bjorklund and Gustafsson, 2005; Guglielmi et al., 2004), most of which are conserved in fungi, plants, and animals (Boube et al., 2002). The overall structural organization of Mediator is conserved in *S. cerevisiae*, *Schizosaccharomyces pombe*, and mammals with four modules termed the head, the middle, the tail, and the Cdk8 module (Dotson et al., 2000; Elmlund et al., 2006). Numerous transcription activators interact with the tail module possibly to recruit the complex to promoters (Bjorklund and Gustafsson, 2005). The Cdk8 module is thought to have a mostly repressive role (Bjorklund and Gustafsson, 2005). The middle and head modules surround Pol II and interact with some of its subunits (Davis et al., 2002). The head module is essential for its interaction with Pol II and for basal and activated transcription (Takagi et al., 2006; Takagi and Kornberg, 2006).

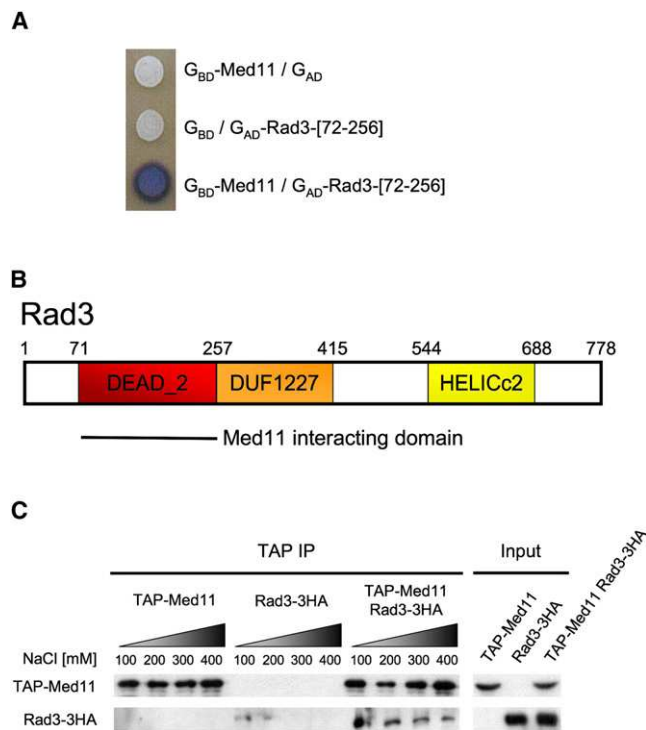
The order of recruitment of various components involved in PIC formation was investigated in vivo in *S. cerevisiae* (Bhoite et al., 2001; Bryant and Ptashne, 2003; Cosma et al., 1999). In this organism, Mediator is recruited before Pol II and the GTFs, suggesting that it is required for their binding. In addition, mutations that abolish the recruitment of Mediator or grossly affect its head module prevent the association of the GTFs (Bhaumik et al., 2004; Takagi et al., 2006). However, current studies do not distinguish whether the GTFs and Pol II are assembled in the PIC in vivo in a stepwise fashion as they are in vitro and how Mediator might stimulate this process. Biochemical experiments have indicated that the establishment of the PIC and the first round of transcription on templates differed from the subsequent ones. Namely, it was observed that Mediator and some GTFs, among which are TFIH and TFIIE, remained bound to the promoter after transcription initiation by Pol II, thereby facilitating reinitiation (Yudkovsky et al., 2000).

In this work, we have found a direct interaction between a head Mediator subunit and a TFIH core subunit. This interaction is important for the association of the TFIK module of TFIH to promoters. In vivo, recruitment of TFIH and TFIIE to the PIC can occur independently of Pol II in a Mediator-dependent fashion. Based on our observations, we propose that several pathways for the assembly of GTFs leading to PIC formation coexist in vivo.

## RESULTS

### Med11 Mediator Subunit Interacts with Rad3 Subunit of TFIH

To better understand Mediator organization and function, we screened its subunits for protein-protein interactions in the two-hybrid system with a yeast DNA fragments library (Fro-mont-Racine et al., 1997). Interactions between Mediator sub-



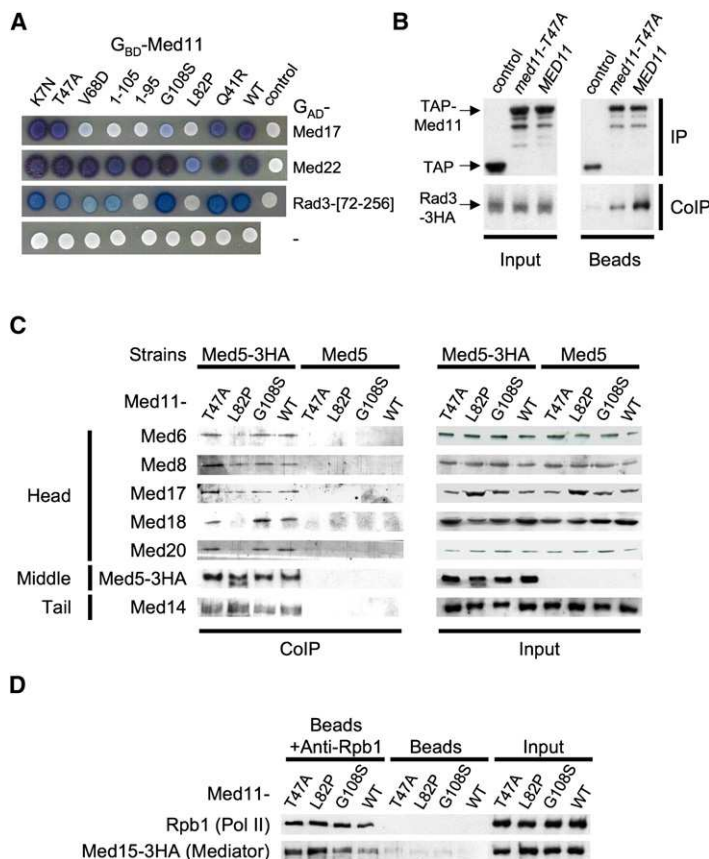
**Figure 1. Med11 Mediator Subunit Interacts with Rad3 Helicase of TFIH**

(A) Two-hybrid interaction between Med11 and Rad3. Two-hybrid assays were performed in Y190 strains transformed by pGBT9, expressing Gal4 DNA-binding domain (GBD) or pGBT9-Med11 ( $G_{BD}$ -Med11) and pACTII, expressing the Gal4 activating domain ( $G_{AD}$ ) or pACTII-Rad3-[72–256] ( $G_{AD}$ -Rad3-[72–256]). Interaction activated a *lacZ* reporter gene that was revealed by a blue color in an X-gal overlay assay (Werner et al., 1993).

(B) Rad3 conserved domain organization according to PFAM database ([www.sanger.ac.uk/Software/Pfam/](http://www.sanger.ac.uk/Software/Pfam/)). A line indicates the Rad3 interaction domain with Med11.

(C) Coimmunoprecipitation of Med11 and Rad3. Proteins from strains overexpressing TAP-Med11, Rad3-3HA, or both proteins were immunoprecipitated with magnetic beads coupled to IgG. The IPs were performed in buffers containing 100 to 400 mM NaCl. After immunoprecipitation, the proteins were revealed by western blotting using 12CA5 or PAP antibodies, binding to HA or TAP tag, respectively.

units have been previously reported (Guglielmi et al., 2004; Kang et al., 2001; Takagi et al., 2006). Additional interactions between Mediator subunits and proteins belonging to other nuclear complexes were found in this screen. In particular, Med11, a 131 amino acid subunit of Mediator head (Gustafsson et al., 1998), was found to interact with Rad3, one of the two helicases of TFIH (Figure 1A), suggesting a role for Mediator in TFIH recruitment. Med11 is essential for yeast survival and is a conserved protein in the eukaryotic kingdom (Backstrom et al., 2007; Boube et al., 2002; Sato et al., 2003). It is important for Mediator head module organization (Takagi et al., 2006). In addition to its interaction with Rad3, Med11 binds Med17 and Med22 in the two-hybrid system (Guglielmi et al., 2004). The 72–256 fragment of Rad3 that interacted with Med11 corresponded to its DEAD\_2 domain (according to the PFAM database; Figure 1B), a conserved region present in a large number of DNA-binding



**Figure 2. Med11 Mutations Affect the Interactions with Its Partners**

(A) Effect of the mutations on two-hybrid interactions between Med11 and its partners. Two-hybrid assays were performed in Y190 strains transformed by pGBT9 or pGBT9-Med11 mutants (G<sub>BD</sub> or G<sub>BD</sub>-Med11), pACTII, pACTII-Med17, pACTII-Med22, or pACTII-Rad3-[72-256] (G<sub>AD</sub>, G<sub>AD</sub>-Med17, G<sub>AD</sub>-Med22 or G<sub>AD</sub>-Rad3-[72-256]). The X-gal agarose assay was performed as in Figure 1. An interaction between Med11 or its mutants and Med17, Med22, or Rad3-[72-256] is indicated by the blue color. A paler shade of blue compared to wild-type indicates a decrease in the strength of interaction between a Med11 mutant protein and its partner.

(B) Med11-T47A mutation decreases Rad3 coimmunoprecipitation with Mediator. Strains expressing TAP, TAP-Med11 or TAP-Med11-T47A, and Rad3-3HA were grown to exponential phase at 30°C. Protein extracts of these strains were immunoprecipitated in a buffer containing 400 mM NaCl with magnetic beads coupled to mouse IgG (Invitrogen). After immunoprecipitation, the proteins were revealed by western blotting using 12CA5 or PAP antibodies.

(C) Mediator integrity revealed by CoIP in mutant strains. Wild-type and mutant strains were grown 45 min at 37°C. Med5-3HA was immunoprecipitated from cell extracts using magnetic beads coupled to anti-HA antibodies. Coimmunoprecipitations of Mediator subunits were revealed by western blotting using anti-HA, -Med6, -Med8, -Med14, -Med17, -Med18, or -Med20 antibodies.

(D) Mediator and Pol II coimmunoprecipitated in Med11 mutant strains. Wild-type and mutant strains were grown 45 min at 37°C. Pol II was immunoprecipitated from cell extracts using magnetic beads coupled to an anti-Rpb1 antibody (8WG16). Negative controls were magnetic beads without antibody. After immunoprecipitation, the coimmunoprecipitated proteins were revealed by western blotting using anti-Rpb1 or anti-HA (12CA5) antibodies.

helicases. Members of this family include proteins of eukaryotic, bacterial, and archaeal origins. To confirm the interaction between the two proteins, TAP-Med11 and Rad3-3HA were overexpressed on a 2  $\mu$  plasmid under the control of the *PGK1* promoter and immunoprecipitated from crude extracts using magnetic beads coupled to IgG. Med11 fused to the TAP tag (Rigaut et al., 1999) coimmunoprecipitated (CoIP) with Rad3 fused to a 3HA tag (Figure 1C). The interaction was resistant to 400 mM NaCl.

### Mutations of Med11 Alter Specific Interactions with Its Partners

To investigate the role of Med11 in PIC organization and to determine the importance of the interaction between Mediator and TFIIH during the initiation step of Pol II transcription, mutations affecting specifically one contact at a time between Med11 and its partners were selected. We mutagenized *MED11* by error-prone PCR and introduced the mutated fragments in pV204 vector bearing the *TRP1* marker by gap repair in the yCE001 strain in which the *med11* deletion was complemented by wild-type *MED11* borne on a *URA3* vector. After chasing the wild-type *MED11* allele on 5-FOA, colonies growing at 30°C were replica plated at 37°C to identify thermosensitive colonies. As an alternative source of *med11* mutants, we separated the three mutations present in *med11-1* (Han et al., 1999). In total, we obtained eight independent thermosensitive mutations distributed along the protein, three in the N terminus

of Med11, two in the middle of the protein, and one in its C terminus (Figure S1). In addition, two truncation mutants that eliminated part of the C terminus of Med11 were selected. The phenotype of these mutants indicated that this portion of Med11, although important for its function, was not required for cell survival. The growth of *med11-L82P* and the two truncation mutants were even affected at 30°C. These mutants had the strongest growth phenotype at 37°C. The growth of all other mutants, except *med11-Q41R* that was only marginally affected, was strongly reduced at 37°C.

We then tested the ability of Med11 mutant subunits to interact with Med17, Med22, and Rad3-[72-256] in two-hybrid assays (Figure 2A). Except for Med11-Q41R, each mutant showed a decreased interaction with at least one partner. The two truncation mutations and the three point mutations in the C-terminal half of Med11 (Med11-V68D, -L82P, and -G108S) affected the interaction with Med17. This observation suggested that the Med11 C-terminal half is required to bind Med17. Med11-L82P was the only mutant which impaired Med11 association with Med22. In addition, six mutants no longer interacted with Rad3 interaction. Med11-[1-95] and Med11-L82P were strongly affected, while four mutations (Med11-K7N, -T47A, -V68D, and -[1-105]) had a milder effect as quantified by a  $\beta$ -galactosidase assay for Med11-T47A (Figure S2). A coimmunoprecipitation (CoIP) assay was performed to confirm that Med11-T47A was affected in Rad3 interaction. We immunoprecipitated TAP-tagged wild-type Med11 and Med11-T47A in extracts from strains



expressing Rad3-3HA. In agreement with the two-hybrid assay, the *med11-T47A* mutation impaired Rad3 coimmunoprecipitation with Med11 (Figure 2B). The other mutants were unaffected in their interaction with Rad3 (data not shown). Moreover, we verified by  $\beta$ -galactosidase and coimmunoprecipitation assays that Med11-T47A mutation did not alter the interaction of Med11 with its two other partners, Med17 and Med22 (Figure S2).

Three mutants were selected for further study: *med11-G108S* and *med11-T47A*, since they are affected for only one interaction (with Med17 and Rad3, respectively), and *med11-L82P* that was affected in its interactions with all of Med11 partners.

### Effect of *med11* Mutations on Mediator Stability, Mediator Interaction with Pol II, and mRNA Levels

Since Mediator is required for the transcription of nearly all yeast class II genes, we tested if the level of Pol II, GTFs, or Mediator itself could be affected in *med11* mutants at the restrictive temperature. The levels of subunits of complexes belonging to the Pol II transcription machinery were measured by western blotting using subunit-specific antibodies or a 3HA epitope in tagged strains. Protein levels of selected subunits of Mediator (Med6, Med8, TAP-Med11, Med17, Med18 [head]; Med5-3HA, Med7-3HA, Med21-3HA [middle]; Med14, Med15-3HA [tail]; and Cdk8-3HA [Cdk8]), TFIID (Ssl1-3HA and Rad3-3HA [cTFIID]; Ccl1-3HA and Kin28-3HA [TFIID]), TFIIE (Tfa1-3HA), TBP, and Pol II (Rpb1) did not change, indicating that the *med11* mutations did not affect the level of the Pol II transcription machinery (Figure S3).

The effect of the mutations on Mediator integrity was verified by CoIP between subunits of the head, middle, and tail modules (Figure 2C). In *med11-T47A*, *med11-G108S*, and wild-type strains, we observed that Mediator head and tail module subunits coimmunoprecipitated with the Med5 middle module subunit. In contrast, the *med11-L82P* mutation destabilized the association of Med6, Med18, and Med20 with the head but did not affect the head binding of Med8 and Med17. Since Mediator is associated with Pol II in soluble cell extracts (Kim et al., 1994), we tested if the mutations could also alter this association. Pol II was immunoprecipitated with anti-CTD antibodies in extracts from wild-type or *med11* mutant strains expressing Med15-3HA. As shown in Figure 2D, the mutations did not significantly affect Mediator association with Pol II.

We investigated the effect of *med11* mutations after transfer to 37°C on the mRNA abundance of five genes (*ADH1*, *PYK1*, *GLT1*, *MET17*, and *VTC3*) that are expressed at different levels. All mutants affected the level of *ADH1*, *PYK1*, *VTC3*, and *MET17* mRNAs, but only marginally affected that of *GLT1*. As expected from its strong phenotype, *med11-L82P* had the most pronounced defect (Figure S4).

### *med11* Mutations Differentially Affected the Promoter Occupancy of Pol II, TFIIE and TFIID Core, and Kinase Modules

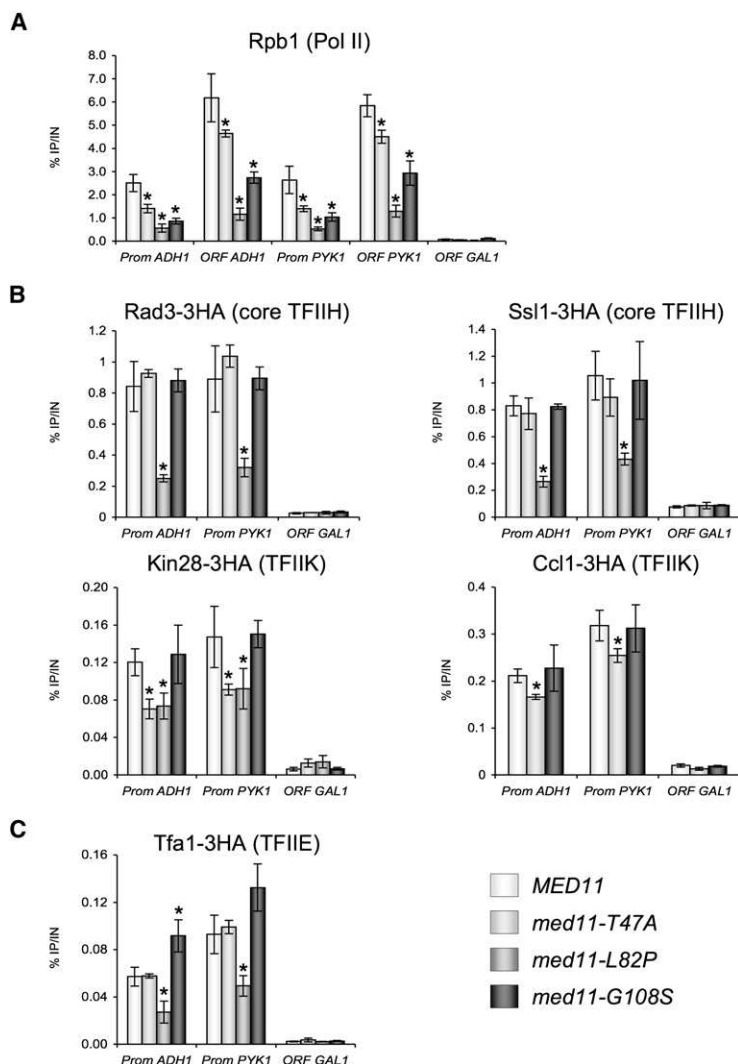
The effect of *med11* mutations on the gene occupancy of Mediator and Pol II or the GTFs on promoters was examined. We immunoprecipitated the chromatin (ChIP) prepared from mutant strains grown in YPD-rich medium for 45 min at the restrictive temperature (37°C) to investigate PIC formation. Chromatin

bound to Mediator was immunoprecipitated from Med5-3HA and Cdk8-3HA strains. The association of Mediator, as seen by ChIP of Med5 and Cdk8, with the promoters of *ADH1*, *PYK1*, *GLT1*, *VTC3*, and *MET17* remained unchanged (Figures S5A and S5B). TBP occupancy in *med11* mutants and the wild-type strains were similar (Figure S5C).

Rpb1-associated chromatin was immunoprecipitated with 8WG16 anti-CTD antibody (Figure 3A and Figure S5D). The *med11-T47A* mutation led to a slight decrease in Rpb1 occupancy (23%–49% depending on the promoter and ORF) while *med11-G108S* led to a larger defect (55%–73% decrease). The *med11-L82P* mutant had a more pronounced defect on Pol II occupancy (77%–83% decrease) consistent with its severe growth phenotype. In vitro and in the absence of Mediator, TFIID, together with TFIIE, is the last GTF to enter the PIC (Buratowski et al., 1989). Since Mediator and TFIID contact each other via the Med11-Rad3 interaction, we wondered if *med11* mutations could impair TFIID occupancy on promoters in vivo. Two subunits of cTFIID (Rad3 and Ssl1) and of its CTD kinase module TFIID (Kin28 and Ccl1) were 3HA-tagged for ChIP experiments. Ccl1 could not be tagged in the *med11-L82P* strain since the addition of the tag was lethal in this genetic background (Figure S6). The *med11-L82P* mutant showed the most pronounced defect since the association of cTFIID and TFIID with promoters was impaired drastically (Figure 3B and Figure S5E). Unexpectedly, *med11-G108S* did not affect promoter occupancy by cTFIID or TFIID, indicating that Pol II association with promoters is not a prerequisite for TFIID recruitment in vivo. The *med11-T47A* mutation did not lead to lower occupancy of cTFIID but had an effect as large as that of *med11-L82P* on TFIID binding to some promoters, indicating that the two TFIID subcomplexes were recruited or stabilized independently of each other by Mediator. Since TFIIE and TFIID enter the PIC together in vitro, we investigated TFIIE occupancy in *med11* mutants. Interestingly, neither the *med11-T47A* nor the *med11-G108S* mutation affected Tfa1-3HA binding (Figure 3C and Figure S5F). This indicated that Mediator is the prime determinant for the recruitment of TFIIE to the PIC in vivo.

### Effect of the *med11-T47A* Mutation Genome-wide on TFIID Binding

*med11-T47A* affected the association of Kin28 with *ADH1* and *PYK1*, but not *GLT1*, *VTC3*, or *MET17*. This led to the question of how many genes in the yeast genome had reduced binding for Kin28. We immunoprecipitated chromatin bound to Kin28 from the wild-type or a *med11-T47A* mutant followed by DNA hybridization on microarrays (ChIP-chip) that allowed us to resolve the position of Kin28 on chromatin to within 250 nt (see the Experimental Procedures). Comparing Kin28 occupancy in the two strains, we found 787 probes corresponding to 665 genes where the binding of Kin28 was reduced 2-fold or more in the mutant (Figure 4A). This criterion was rather conservative and probably leads to an underestimation of the number of genes where the binding of TFIID was affected. Similar ChIP-chip experiments were used to investigate the effect of the *med11-T47A* mutation on the genome-wide distribution of Rad3 and Med5. Interestingly, most of the genes that were affected for Kin28 binding were not affected for Rad3 binding. As described



**Figure 3. Effect of *med11* Mutations on Pol II, TFIIK, cTFIIH, and TFIIK Occupancy**

Immunoprecipitated DNA from ChIP experiments was amplified with primers as indicated in the Supplemental Experimental Procedures. *GAL1* open reading frame (ORF) was used as a negative control since its transcription is repressed in YPD. Values represent the average of three independent experiments. Error bars indicate the standard deviation and an asterisk indicates a significant difference between the wild-type and the mutant at  $p$  value  $<0.05$  in a Student's  $t$  test.

(A) Pol II occupancy. Cells were grown on YPD to  $OD_{0.6}$  at 600 nm, then transferred for 45 min to 37°C. Standard ChIP assays were performed on chromatin prepared from WT, *med11-T47A*, *med11-L82P*, and *med11G108S* strains using an antibody against Rpb1 CTD (8WG16).

(B) cTFIIH and TFIIK subunits occupancy. Strains expressing Rad3-3HA, Ssl1-3HA, Kin28-3HA, or Ccl1-3HA in wild-type or mutant backgrounds were grown for 45 min at 37°C. Standard ChIP assays were performed using anti-HA antibody (12CA5).

(C) TFIIE occupancy. Strains expressing Tfa1-3HA in wild-type or mutant backgrounds were grown for 45 min at 37°C. Standard ChIP assays were performed using anti-HA antibody (12CA5).

significantly affect total Pol II binding, the slope of the correlation line being equal to 1.03. Interestingly, the slope of the regression line between the enrichments of S5P Pol II in the wild-type and the mutant decreased to 0.79, indicating that S5 CTD phosphorylation levels were lower in the mutant.

To confirm the genome-wide phosphorylation defect, we performed ChIP experiments on three selected genes to look at the distribution of total or phosphorylated Pol II. *HST2* was chosen because the ChIP-chip analysis indicated that TFIIK occupancy was reduced. *ADH1* and *PYK1* occupancies were also analyzed since, even though genome-wide analysis did not point to a significant defect in S5P Pol II occupancy for these genes, our ChIP experiments indicated that the level of TFIIK was re-

duced. As seen on Figure 5B, *med11-T47A* had a more pronounced effect on the levels of S5P Pol II than on that of total Pol II population on all three genes as expected from the TFIIK occupancy defect.

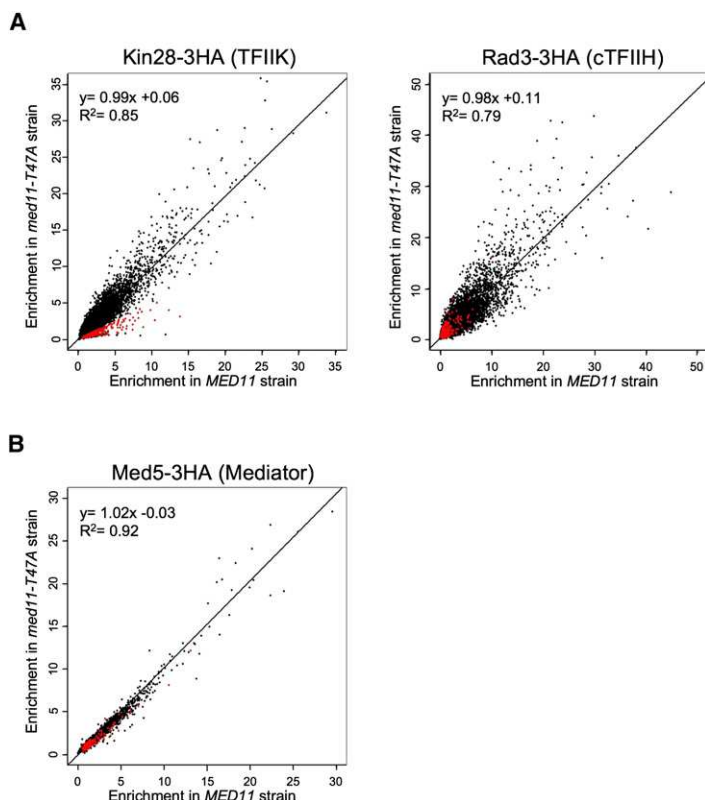
### The Med11-Rad3 Interaction Stimulates CTD Phosphorylation In Vivo

Phosphorylation of Serine 5 of Pol II CTD by TFIIK is an important step for transcription initiation (Akoulitchiev et al., 1995) and is activated by Mediator (Guidi et al., 2004; Nair et al., 2005). Since we found that *med11-T47A* had a defect in TFIIK occupancy on numerous promoters, it was expected that the mutation would lead to decreased CTD phosphorylation on such genes. An antibody (H14) directed against the CTD phosphorylated on serine 5 (S5P Pol II) was used to look at its genome-wide occupancy in a mutant and in a wild-type strain. A similar experiment was performed to analyze the genome-wide distribution of total Pol II population in the two strains (Figure 5A). As predicted from the ChIP experiments shown in Figure 3A, *med11-T47A* did not

above, specific ChIP experiments showed that *med11-T47A* had no effect on Med5 association. This observation was confirmed genome-wide by a ChIP-chip experiment (Figure 4B). Examples of occupancy profiles of Kin28, Rad3, and Med5 proteins on specific genes are shown in Figure S7.

We also measured total Pol II and S5P Pol II occupancies on *ADH1* and *PYK1* in *med11-L82P*- and *med11-G108S*-carrying strains. The levels of both forms of Pol II decreased in both mutant backgrounds, but the ratio of S5P Pol II to total Pol II remained unchanged (data not shown).

In addition to defective TFIIK association in PIC, *med11-T47A* might have also affected the CTD kinase stimulatory activity of Mediator. To test this possibility, we set up a CTD phosphorylation assay containing TFIIH, recombinant GST-CTD, labeled ATP, and wild-type or mutant Mediator. Cellular extracts in which Med5 was not tagged in wild-type or mutant backgrounds were subjected to the same immunopurification procedure and used as a negative control in the CTD phosphorylation assays (Figure S8). Phosphorimager quantification of labeled GST-CTD indicated that stimulation of TFIIH kinase activity was identical for the *med11-T47A* and the wild-type Mediator. Taken together,



**Figure 4. Genome-wide Analysis Reveals that the *med11-T47A* Mutation Specifically Affects TFIIF Occupancy**

(A) Genome-wide analysis of TFIIF modules occupancy. ChIP-chip analysis was performed on Kin28-3HA (TFIIF) and Rad3-3HA (cTFIIF) in *med11-T47A* and wild-type strains with 12CA5 antibody. Wild-type or mutant cells were grown as in Figure 3. The graphic represents the correlation of Kin28-3HA and Rad3-3HA in *med11-T47A* and the wild-type strains. The red dots correspond to the yeast genome part that presents a decrease superior to 2-fold of TFIIF occupancy. A linear regression and its equation are indicated.

(B) Genome-wide analysis of Mediator occupancy. ChIP-chip analysis was performed on Med5-3HA in *med11-T47A* and wild-type strains with 12CA5 antibody. Wild-type or mutant cells were grown as in Figure 3. The graphic represents the correlation of Med5-3HA binding in *med11-T47A* and the wild-type strains. A linear regression and its equation are indicated.

## DISCUSSION

### Mediator Stimulates TFIIF Association to Promoters In Vivo through Direct Interaction

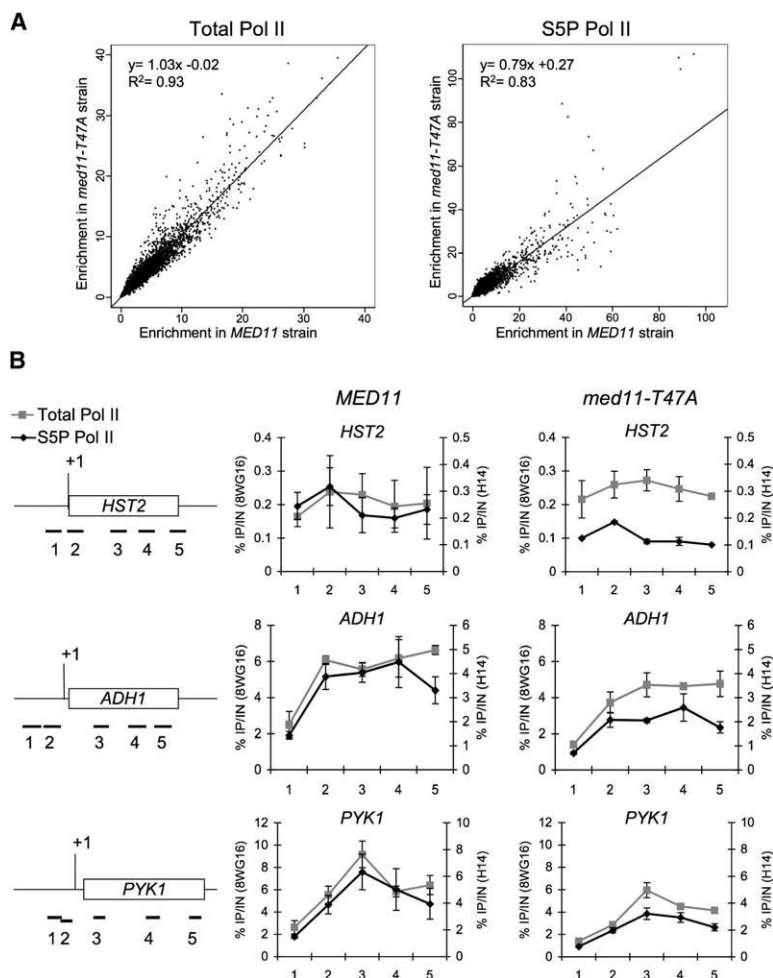
The finding that Med11 interacted with Rad3, both in the two-hybrid system and in coimmunoprecipitation experiments, suggested that Mediator and TFIIF might functionally interact. A recent genetic interaction map between genes encoding subunits of complexes involved in transcription also suggested a connection between the Mediator head and TFIIF (Collins et al., 2007).

To investigate the nature of the link between Mediator and TFIIF, we selected conditional Med11 mutations and investigated their effects on the binding of Mediator itself, GTFs, and Pol II to promoters. The *med11-G108S* mutation impaired the association of Pol II with the promoters that were tested but did not impair association with TFIIE, TFIIF, Mediator itself, and TBP. Another mutation, *med11-L82P*, did lead to decreased occupancy of TFIIF, TFIIE, and Pol II, but not of Mediator or TBP. These observations point to a functional role for Mediator in TFIIF and TFIIE recruitment to promoters and can be interpreted in two ways that are not mutually exclusive. The first interpretation would be that Mediator is important for the recruitment of TFIIF and TFIIE in PIC formation independently of Pol II. A second possibility is that the *med11-G108S* mutation blocks the capture of Pol II by the scaffold and prevents reinitiation. To determine if complexes containing Mediator, TFIIF, and TFIIE in the *med11-G108S* mutant are representative of PIC intermediates, we identified promoters in the wild-type to which TFIIF and Mediator were bound but Pol II was absent. In addition, all 16 centromeres of *S. cerevisiae* were bound by Mediator and TFIIF, but not by Pol II. At present, we do not know the significance of this latter observation. Nevertheless, since the promoters were transcriptionally silent, as were the centromeres, Mediator-TFIIF complexes are more likely to represent PIC intermediates that have not been activated rather than reinitiation scaffolds. PIC intermediates containing TFIIF and TFIIE, but not Pol II, are unexpected in the light of previous results showing that Pol II association with promoters is required for the recruitment of TFIIE and TFIIF in vitro (Buratowski et al., 1989). In this previous work, the

these observations demonstrated that the decreased S5 phosphorylation of CTD in *med11-T47A* is due to the decreased occupancy of TFIIF on promoters as a result of a weaker interaction with Rad3.

### Mediator and TFIIF Can Occupy Promoters and Centromeres in the Absence of Pol II

Since the *med11-G108S* mutant decreased Pol II occupancy on the five genes that were examined without affecting Mediator, TFIIF, or TFIIE binding, Mediator could have been an important determinant for TFIIF recruitment to promoters. If this was the case, one would expect to find genes that are bound by Mediator and TFIIF before transcription activation in the wild-type. Genome-wide location analyses of Rpb1 (Figure 5), Med5, Rad3, and Kin28 (Figure 4) in a wild-type strain allowed us to identify genes that were specifically enriched for Mediator and TFIIF, but not for Pol II. Examples of genes displaying this specific behavior are shown in Figure 6. Unexpectedly, we also found that all 16 centromeres were highly enriched for both Mediator and TFIIF, but not for Pol II. Centromeres are not thought to be transcribed in *S. cerevisiae*. Nevertheless, one could argue that the presence of Mediator and TFIIF is due to the proximity of genes that are transcribed divergently from the centromere. This hypothesis cannot hold for chromosomes 3, 4, 9, 12, 13, and 14 where the genes located on either side of the centromere are transcribed convergently, as shown, for example, in Figure 6 for centromere 14. These observations demonstrate that TFIIF can be recruited independently of Pol II.



**Figure 5. *med11-T47A* Mutation Affects Pol II Phosphorylation**

(A) ChIP-chip analysis of total Pol II and S5P Pol II in *med11-T47A* and wild-type strains. Wild-type or mutant cells were grown as described in Figure 3. ChIP-chip assays were performed using an antibody against total CTD (8WG16) or against CTD phosphorylated on serine 5 (H14). The graphics represent the correlation between Pol II and S5P Pol II occupancy in *med11-T47A* and the wild-type strains. A linear regression was performed, and its equation is indicated.

(B) Profiles of total and S5P Pol II on *HST2*, *ADH1*, and *PYK1* in wild-type and *med11-T47A* strains. On the left, schematic promoter and coding region organization of *HST2*, *ADH1*, and *PYK1* are presented. The location of the PCR fragments amplified in ChIP analysis is indicated. The open reading frame (ORF) is indicated as an open box. The vertical line indicates the position of the transcription initiation site. Wild-type or mutant cells were grown as in Figure 3. Standard ChIP assays were performed using an antibody against total (8WG16) or S5P Pol II (H14). Values represent the average of three independent experiments. Error bars indicate the standard deviation.

chromatin binding of cTFIIH, to which Rad3 belongs, but diminished binding of TFIIK to about 10% of the yeast genes. Since the *med11-T47A* mutation had only a mild effect on the interaction of Mediator with Rad3, the proportion of yeast genes that depend on Med11 for TFIIK association is probably underestimated. A role for Rad3 in stabilizing TFIIK is consistent with the bridging role of Rad3/XPD between the two modules of TFIIH (Drapkin et al., 1996; Reardon et al., 1996) and with the role of Rad3/XPD in facilitating optimal transcription by anchoring the CAK subcomplex to cTFIIH (Tirode et al., 1999).

pathway leading to PIC formation was investigated biochemically using a system that, in fact, lacked Mediator, which had not been discovered at that time. This might explain the discrepancy between the two studies.

We do not exclude the possibility that Mediator in general, and Med11 in particular, plays a role in the stabilization of reinitiation scaffolds in vivo. However, to the best of our knowledge, the existence of scaffolds has never been demonstrated in vivo probably because experimental methods to address this question are particularly difficult to set up.

#### Evidence that Mediator Stimulates Serine 5 Phosphorylation of the CTD by Stabilizing TFIIK on Promoters

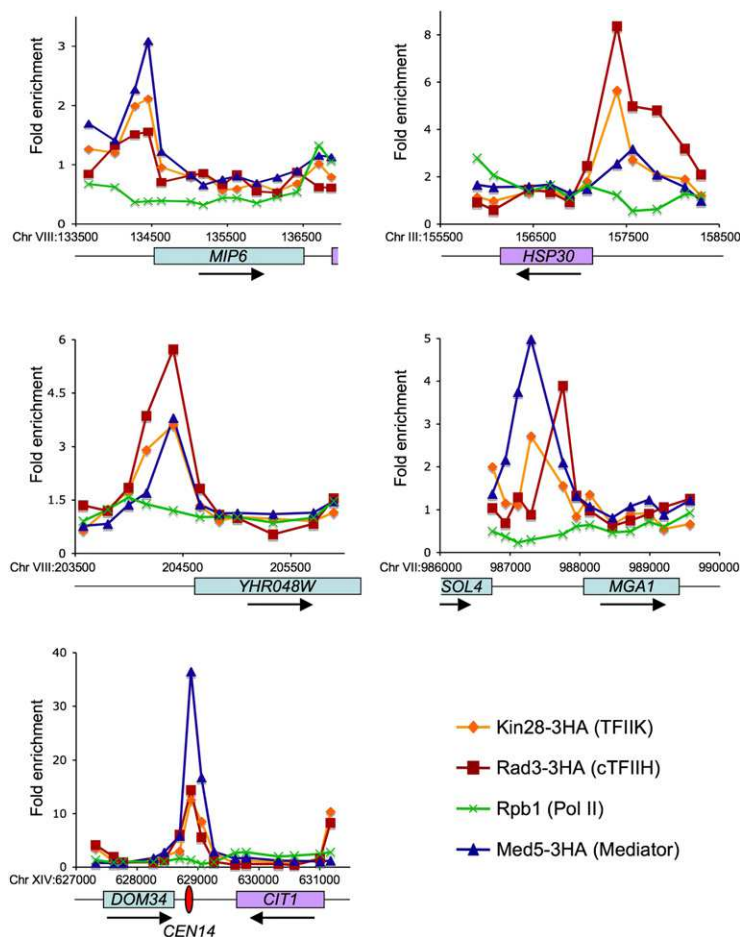
TFIIH controls promoter clearance by Pol II through its helicase activity that is required for DNA strand separation and through its CTD kinase activity that is stimulated by Mediator (Egly, 2001; Kim et al., 1994). Our experiments strongly suggested that the direct interaction between the Med11 subunit of Mediator and the Rad3 subunit of TFIIH via its DEAD\_2 helicase domain plays an important role in the stimulation of TFIIH kinase activity. A mutation, *med11-T47A*, that selectively decreased the interaction between Rad3 and Med11 was identified. It did not decrease

By genome-wide location analysis and ChIP, we observed that the amount of phosphorylated Pol II bound to chromatin decreased in *med11-T47A* compared to wild-type, whereas total Pol II levels remained largely unaffected. This could result from either the decreased association of TFIIK with promoters or from lower kinase activity or both. In vitro assays indicated that the mutation did not impair the kinase stimulatory activity of Mediator, suggesting that in *med11-T47A*, the decrease in CTD phosphorylation is accounted for by the decreased recruitment of TFIIK.

In the two other *med11* mutants that were studied, the amount of total and phosphosphorylated Pol II decreased in the same proportion on *ADH1* and *PYK1* genes (data not shown). The *med11-G108S* mutation did not affect TFIIK binding to promoters, which meant that Pol II phosphorylation was normal. The *med11-L82P* mutation reduced TFIIK binding by about 2-fold, while Pol II binding diminished by at least 5-fold. Since the ratio of Pol II to S5P Pol II did not change, we suggest that TFIIK was not limiting for Pol II phosphorylation as a consequence of the increased TFIIK to Pol II ratio.

The DEAD\_2 domain of TFIIH is highly conserved in eukaryotes. Some mutations in the DEAD\_2 domain of XPD, the human homolog of Rad3, lead to the trichothiodystrophy disease





**Figure 6. Mediator, TFIIF, and Pol II Enrichment over Selected Genes in Wild-Type Strain**

Examples of genomic regions bound by TFIIF and Mediator but not Pol II. ChIP-chip assays were performed using antibodies against Rpb1 (8WG16) or against the HA tag (12CA5) in wild-type strain grown as in Figure 3. The graphics present binding ratios of Med5-3HA (Mediator), Kin28-3HA (TFIIF), Rad3-3HA (cTFIIF), and Rpb1 (Pol II).

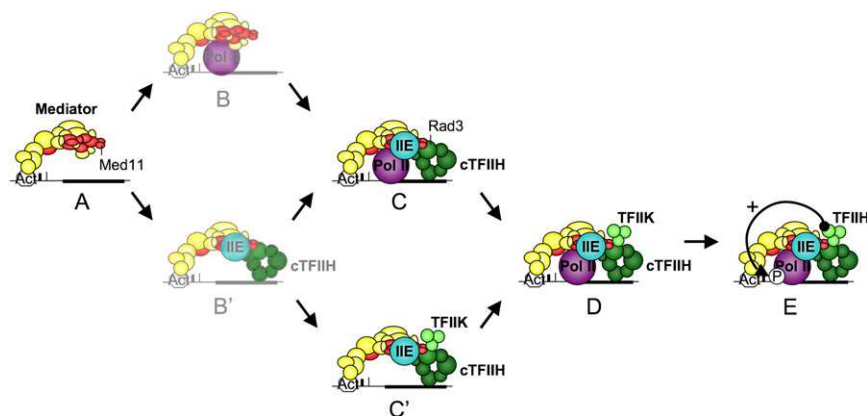
may be that Mediator could play a role in transcription-coupled repair by stimulating TFIIF recruitment following DNA damage. We found that none of the *med11* mutations increased the UV sensitivity of the mutant strains (data not shown), indicating that the loading of TFIIF at promoters stimulated by Mediator is not required for repair and suggesting that TFIIF can access damaged sites directly.

#### A Branched Model for Transcription Initiation

Based on our results, we propose a branched model of PIC formation shown on Figure 7. Med11-T47A diminished the association of TFIIF to promoters without affecting that of Pol II, TFIIE, or cTFIIF. Conversely, Med11-G108S affected Pol II association, but not that of TFIIE, cTFIIF, or TFIIF, suggesting the existence of two PIC intermediate complexes that would contain Mediator•Pol II•TFIIE•cTFIIF (Figure 7C) and Mediator•TFIIE•TFIIF (Figure 7C'). The ChIP-chip data demonstrated the existence of such complexes on inactive promoters and centromeres. Mediator•Pol II•TFIIE•cTFIIF and Mediator•TFIIE•TFIIF complexes could lead to a full PIC (Figure 7D) in a single step by the recruitment of a single factor: TFIIF in the first case, and Pol II in the second. Two hypothetical intermediates could lead to Mediator•Pol II•TFIIE•cTFIIF formation. The first one would entail a complex consisting of Mediator and Pol II (Figure 7B) that has not been observed in vivo but can be proposed on the basis of the association of Pol II and Mediator in cell extracts (Kim et al., 1994). The second one would consist of Mediator•TFIIE•cTFIIF as suggested by our observation that TFIIF is not always present simultaneously with cTFIIF

characterized by defects in basal and activated transcription (Dubaele et al., 2003). In the light of our results, it is possible that these XPD mutations are defective in CTD serine 5 phosphorylation because of the poor interaction between TFIIF and Mediator.

TFIIF plays a critical role in NER, in general, and transcription coupled repair (TCR), in particular (Laine and Egly, 2006). Since Mediator is required for TFIIF association with promoters, it



**Figure 7. Branched Pathway Model of PIC Formation**

(A–E) The steps leading to PIC formation are depicted in the cartoon. Arrows connect the pre-initiation intermediates in a branched pathway leading to PIC formation and Pol II phosphorylation. The intermediates that have been observed are indicated by solid colors. The intermediates that were inferred are indicated by the translucent colors.

The color code used is as follows: red, essential Mediator subunits; yellow, dispensable Mediator subunits; dark green, cTFIIF; light green, TFIIF; blue, TFIIE; violet, Pol II.

(Figure 7B'). Interestingly, this intermediate could also lead to the formation of the Mediator•TFII E•TFIH complex by the recruitment of TFIK.

In conclusion, Mediator appears to play a critical role in PIC assembly at the level of TFIH and TFII E recruitment. Several pathways leading to PIC formation may coexist with several intermediates that might be more or less abundant depending on specific promoters. The various pathways leading to PIC formation could be differentially regulated, expanding the repertoire of mechanisms for regulating gene transcription.

## EXPERIMENTAL PROCEDURES

### Oligonucleotides

The sequence of the oligonucleotides used in this study can be found in Table S1.

### Chromatin Immunoprecipitation and Genome-wide Location Analysis

Crosslinked chromatin was prepared essentially as described previously (Kuras et al., 2003). More details can be found in the Supplemental Data.

For the ChIP-chip experiments, ligation-mediated PCR was done as described previously (Ren et al., 2000), except that amino-allyl conjugated dUTP (150  $\mu$ M final) was used, and only 30 cycles of PCR were performed. The PCR products were purified using a Microcon YM-30 filter. Amino-allyl modified DNA was recovered with 20  $\mu$ l of H<sub>2</sub>O and the DNA was lyophilized. DNA was labeled as previously described (Harismendy et al., 2003), except that unincorporated dyes were removed using a QIAquick PCR Purification Kit (QIAGEN). Labeled DNA was recovered with 100  $\mu$ l of buffer EB, and ethanol was precipitated. Hybridization and washing conditions were as described previously (Lee et al., 2006). Microarrays were obtained from Agilent Technologies (G4493A) and feature 41,776 60-mer oligonucleotide probe spots with an average density of one probe each 266 bp of the yeast genome. Images of Cy5 and Cy3 fluorescence intensities were generated using the Genepix 4000B scanner (Molecular Devices) and extracted using GenePix Pro 6 software (Molecular Devices). Experiments were carried out at least in triplicate.

### Data Analysis

The computational data analysis was performed in R using the *limma* package (Smyth et al., 2005) from the bioconductor project ([www.bioconductor.org](http://www.bioconductor.org)). After subtraction of the local background, the data from the two channels were median normalized, and log ratios of signal intensities were generated for each feature. The log ratios were processed by fitting a linear model for each feature, and p values were calculated by performing an empirical Bayes moderated t statistic test. p values were then adjusted for multiple testing by Benjamini and Hochberg false discovery rate (FDR) method.

### Statistical Analyses

Statistical analyses were performed using GraphPad (Graphpad Software Inc.). Differences between mutants and wild-type were compared by the Student's t test and considered statistically significant at the  $p < 0.05$  level of confidence. Data are expressed as the mean  $\pm$  standard deviation.

### ACCESSION NUMBERS

The complete raw data set is available at Array Express (<http://www.ebi.ac.uk/arrayexpress>) under accession number E-MEXP-1502.

### SUPPLEMENTAL DATA

The Supplemental Data include Supplemental Experimental Procedures, eight figures, two tables, and Supplemental References and can be found with this article online at <http://www.molecular.org/cgi/content/full/31/3/337/DC1/>.

## ACKNOWLEDGMENTS

We thank S. Björklund, S. Hahn, R. Kornberg, and A. Weil for antibodies; D. Hermand for strains; and J. Ranish for attempting to analyze the role of the Med11 in vitro. We also thank, B. Guglielmi and J.-M. Egly for discussions and A. Sentenac, S. Un, P. Thuriaux, and C. Mann for critical reading of the manuscript. C.E. was supported by a grant from the Association pour la Recherche sur le Cancer.

Received: July 13, 2007

Revised: February 10, 2008

Accepted: June 1, 2008

Published: August 7, 2008

## REFERENCES

- Akoulitchiev, S., Mäkelä, T.P., Weinberg, R.A., and Reinberg, D. (1995). Requirement for TFIH kinase activity in transcription by RNA polymerase II. *Nature* 377, 557–560.
- Backstrom, S., Elfving, N., Nilsson, R., Wingsle, G., and Björklund, S. (2007). Purification of a plant mediator from *Arabidopsis thaliana* identifies PFT1 as the Med25 subunit. *Mol. Cell* 26, 717–729.
- Baek, H.J., Kang, Y.K., and Roeder, R.G. (2006). Human Mediator enhances basal transcription by facilitating recruitment of transcription factor IIB during preinitiation complex assembly. *J. Biol. Chem.* 281, 15172–15181.
- Bhaumik, S.R., Raha, T., Aiello, D.P., and Green, M.R. (2004). In vivo target of a transcriptional activator revealed by fluorescence resonance energy transfer. *Genes Dev.* 18, 333–343.
- Bhoite, L.T., Yu, Y., and Stillman, D.J. (2001). The Swi5 activator recruits the Mediator complex to the HO promoter without RNA polymerase II. *Genes Dev.* 15, 2457–2469.
- Björklund, S., and Gustafsson, C.M. (2005). The yeast Mediator complex and its regulation. *Trends Biochem. Sci.* 30, 240–244.
- Boube, M., Joulia, L., Cribbs, D.L., and Bourbon, H.M. (2002). Evidence for a mediator of RNA polymerase II transcriptional regulation conserved from yeast to man. *Cell* 110, 143–151.
- Bryant, G.O., and Ptashne, M. (2003). Independent recruitment in vivo by Gal4 of two complexes required for transcription. *Mol. Cell* 11, 1301–1309.
- Buratowski, S., Hahn, S., Guarente, L., and Sharp, P.A. (1989). Five intermediate complexes in transcription initiation by RNA polymerase II. *Cell* 56, 549–561.
- Chang, W.-H., and Kornberg, R.D. (2000). Electron crystal structure of the transcription and DNA repair complex, core TFIH. *Cell* 102, 609–613.
- Collins, S.R., Miller, K.M., Maas, N.L., Roguev, A., Fillingham, J., Chu, C.S., Schuldiner, M., Gebbia, M., Recht, J., Shales, M., et al. (2007). Functional dissection of protein complexes involved in yeast chromosome biology using a genetic interaction map. *Nature* 446, 806–810.
- Conaway, R.C., Sato, S., Tomomori-Sato, C., Yao, T., and Conaway, J.W. (2005). The mammalian Mediator complex and its role in transcriptional regulation. *Trends Biochem. Sci.* 30, 250–255.
- Cosma, M.P., Tanaka, T., and Nasmyth, K. (1999). Ordered recruitment of transcription and chromatin remodeling factors to a cell cycle and developmentally regulated promoter. *Cell* 97, 299–311.
- Davis, J.A., Takagi, Y., Kornberg, R.D., and Asturias, F.A. (2002). Structure of the Yeast RNA Polymerase II Holoenzyme. Mediator Conformation and Polymerase Interaction. *Mol. Cell* 10, 409–415.
- Dotson, M.R., Yuan, C.X., Roeder, R.G., Myers, L.C., Gustafsson, C.M., Jiang, Y.W., Li, Y., Kornberg, R.D., and Asturias, F.J. (2000). Structural organization of yeast and mammalian mediator complexes. *Proc. Natl. Acad. Sci. USA* 97, 14307–14310.
- Drapkin, R., Le Roy, G., Cho, H., Akoulitchiev, S., and Reinberg, D. (1996). Human cyclin-dependent kinase-activating kinase exists in three distinct complexes. *Proc. Natl. Acad. Sci. USA* 93, 6488–6493.

- Dubaele, S., Proietti De Santis, L., Bienstock, R.J., Keriell, A., Stefanini, M., Van Houten, B., and Egly, J.M. (2003). Basal transcription defect discriminates between xeroderma pigmentosum and trichothiodystrophy in XPD patients. *Mol. Cell* 11, 1635–1646.
- Egry, J.M. (2001). The 14th Datta Lecture. TFIIF: from transcription to clinic. *FEBS Lett.* 498, 124–128.
- Elmlund, H., Baraznenok, V., Lindahl, M., Samuelsen, C.O., Koeck, P.J., Holmberg, S., Hebert, H., and Gustafsson, C.M. (2006). The cyclin-dependent kinase 8 module sterically blocks Mediator interactions with RNA polymerase II. *Proc. Natl. Acad. Sci. USA* 103, 15788–15793.
- Feaver, W.J., Svejstrup, J.Q., Henry, N.L., and Kornberg, R.D. (1994). Relationship of CDK-activating kinase and RNA polymerase II CTD kinase TFIIF/TFIIK. *Cell* 79, 1103–1109.
- Fisher, R.P. (2005). Secrets of a double agent: CDK7 in cell-cycle control and transcription. *J. Cell Sci.* 118, 5171–5180.
- Flanagan, P.M., Kelleher, R.J., III, Sayre, M.H., Tschoshner, H., and Kornberg, R.D. (1991). A mediator required for activation of RNA polymerase II transcription in vitro. *Nature* 350, 436–438.
- Fromont-Racine, M., Rain, J.-C., and Legrain, P. (1997). Toward a functional analysis of the yeast genome through exhaustive two-hybrid screens. *Nat. Genet.* 16, 277–282.
- Giglia-Mari, G., Coin, F., Ranish, J.A., Hoogstraten, D., Theil, A., Wijgers, N., Jaspers, N.G., Raams, A., Argentini, M., van der Spek, P.J., et al. (2004). A new, tenth subunit of TFIIF is responsible for the DNA repair syndrome trichothiodystrophy group A. *Nat. Genet.* 36, 714–719.
- Guglielmi, B., van Berkum, N.L., Klapholz, B., Bijma, T., Boube, M., Boschiero, C., Bourbon, H.M., Holstege, F.C., and Werner, M. (2004). A high resolution protein interaction map of the yeast Mediator complex. *Nucleic Acids Res.* 32, 5379–5391.
- Guidi, B.W., Bjornsdottir, G., Hopkins, D.C., Lacomis, L., Erdjument-Bromage, H., Tempst, P., and Myers, L.C. (2004). Mutual targeting of mediator and the TFIIF kinase Kin28. *J. Biol. Chem.* 279, 29114–29120.
- Gustafsson, C.M., Myers, L.C., Beve, J., Späth, H., Lui, M., Erdjument-Bromage, H., Tempst, P., and Kornberg, R.D. (1998). Identification of new mediator subunits in the RNA polymerase II holoenzyme of *Saccharomyces cerevisiae*. *J. Biol. Chem.* 273, 30851–30854.
- Guzder, S.N., Qiu, H., Sommers, C.H., Sung, P., Prakash, L., and Prakash, S. (1994a). DNA repair gene RAD3 of *S. cerevisiae* is essential for transcription by RNA polymerase II. *Nature* 367, 91–94.
- Guzder, S.N., Sung, P., Bailly, V., Prakash, L., and Prakash, S. (1994b). RAD25 is a DNA helicase required for DNA repair and RNA polymerase II transcription. *Nature* 369, 578–581.
- Han, S.J., Lee, Y.C., Gim, B.S., Ryu, G.H., Park, S.J., Lane, W.S., and Kim, Y.J. (1999). Activator-specific requirement of yeast mediator proteins for RNA polymerase II transcriptional activation. *Mol. Cell Biol.* 19, 979–988.
- Harismendy, O., Gendrel, C.G., Soularue, P., Gidrol, X., Sentenac, A., Werner, M., and Lefebvre, O. (2003). Genome-wide location of yeast RNA polymerase III transcription machinery. *EMBO J.* 22, 4738–4747.
- Holstege, F.C., van der Vliet, P.C., and Timmers, H.T.M. (1996). Opening of an RNA polymerase II promoter occurs in two distinct steps and requires the basal transcription factors IIE and IIH. *EMBO J.* 15, 1666–1677.
- Holstege, F.C.P., Jennings, E.G., Wyrick, J.J., Lee, T.I., Hengartner, C.J., Green, M.R., Golub, T.R., Lander, E.S., and Young, R.A. (1998). Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* 95, 717–728.
- Johnson, K.M., Wang, J., Smallwood, A., Arayata, C., and Carey, M. (2002). TFIID and human mediator coactivator complexes assemble cooperatively on promoter DNA. *Genes Dev.* 16, 1852–1863.
- Kang, J.S., Kim, S.H., Hwang, M.S., Han, S.J., Lee, Y.C., and Kim, Y.J. (2001). The structural and functional organization of the yeast mediator complex. *J. Biol. Chem.* 276, 42003–42010.
- Kelleher, R.J., III, Flanagan, P.M., and Kornberg, R.D. (1990). A novel mediator between activator proteins and the RNA polymerase II transcription apparatus. *Cell* 61, 1209–1215.
- Keogh, M.C., Cho, E.J., Podolny, V., and Buratowski, S. (2002). Kin28 is found within TFIIF and a Kin28-Ccl1-Tfb3 trimer complex with differential sensitivities to T-loop phosphorylation. *Mol. Cell Biol.* 22, 1288–1297.
- Kim, Y.-J., Björklund, S., Li, Y., Sayre, M.H., and Kornberg, R.D. (1994). A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. *Cell* 77, 599–608.
- Koleske, A.J., and Young, R.A. (1994). An RNA polymerase II holoenzyme responsive to activators. *Nature* 368, 466–469.
- Kuras, L., Borggreve, T., and Kornberg, R.D. (2003). Association of the Mediator complex with enhancers of active genes. *Proc. Natl. Acad. Sci. USA* 100, 13887–13891.
- Laine, J.P., and Egly, J.M. (2006). When transcription and repair meet: a complex system. *Trends Genet.* 22, 430–436.
- Lee, T.I., Johnstone, S.E., and Young, R.A. (2006). Chromatin immunoprecipitation and microarray-based analysis of protein location. *Nat. Protocols* 1, 729–748.
- Max, T., Sogaard, T.M., and Svejstrup, J.Q. (2007). Hyperphosphorylation of the carboxyterminal repeat domain of RNA polymerase II facilitates dissociation of its complex with mediator. *J. Biol. Chem.* 282, 14113–14120.
- Nair, D., Kim, Y., and Myers, L.C. (2005). Mediator and TFIIF govern carboxy-terminal domain-dependent transcription in yeast extracts. *J. Biol. Chem.* 280, 33739–33748.
- Ranish, J.A., and Hahn, S. (1996). Transcription: basal factors and activation. *Curr. Opin. Genet. Dev.* 6, 151–158.
- Ranish, J.A., Hahn, S., Lu, Y., Yi, E.C., Li, X.J., Eng, J., and Aebersold, R. (2004). Identification of TFB5, a new component of general transcription and DNA repair factor IIH. *Nat. Genet.* 36, 707–713.
- Reardon, J.T., Ge, H., Gibbs, E., Sancar, A., Hurwitz, J., and Pan, Z.Q. (1996). Isolation and characterization of two human transcription factor IIH (TFIIH)-related complexes: ERCC2/CAK and TFIIF. *Proc. Natl. Acad. Sci. USA* 93, 6482–6487.
- Ren, B., Robert, F., Wyrick, J.J., Aparicio, O., Jennings, E.G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E., et al. (2000). Genome-wide location and function of DNA binding proteins. *Science* 290, 2306–2309.
- Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M., and Seraphin, B. (1999). A generic protein purification method for protein complex characterization and proteome exploration. *Nat. Biotechnol.* 17, 1030–1032.
- Sato, S., Tomomori-Sato, C., Banks, C.A., Sorokina, I., Parmely, T.J., Kong, S.E., Jin, J., Cai, Y., Lane, W.S., Brower, C.S., et al. (2003). Identification of mammalian Mediator subunits with similarities to yeast Mediator subunits Srb5, Srb6, Med11, and Rox3. *J. Biol. Chem.* 278, 15123–15127.
- Schultz, P., Fribourg, S., Poterszman, A., Mallouh, V., Moras, D., and Egly, J.M. (2000). Molecular structure of human TFIIF. *Cell* 102, 599–607.
- Smyth, G.K., Michaud, J., and Scott, H.S. (2005). Use of within-array replicate spots for assessing differential expression in microarray experiments. *Bioinformatics* 21, 2067–2075.
- Takagi, Y., and Kornberg, R.D. (2006). Mediator as a general transcription factor. *J. Biol. Chem.* 281, 80–89.
- Takagi, Y., Calero, G., Komori, H., Brown, J.A., Ehrensberger, A.H., Hudmon, A., Asturias, F., and Kornberg, R.D. (2006). Head module control of mediator interactions. *Mol. Cell* 23, 355–364.
- Tirole, F., Busso, D., Coin, F., and Egly, J.M. (1999). Reconstitution of the transcription factor TFIIF: assignment of functions for the three enzymatic subunits, XPB, XPD, and cdk7. *Mol. Cell* 3, 87–95.
- Werner, M., Chaussivert, N., Willis, I.M., and Sentenac, A. (1993). Interaction between a complex of RNA polymerase III subunits and the 70 kDa component of TFIIF. *J. Biol. Chem.* 268, 20721–20724.
- Yudkovsky, N., Ranish, J.A., and Hahn, S. (2000). A transcription reinitiation intermediate that is stabilized by activator. *Nature* 408, 225–229.
- Zurita, M., and Merino, C. (2003). The transcriptional complexity of the TFIIF complex. *Trends Genet.* 19, 578–584.

## Supplemental Data

### Mediator-Dependent Recruitment of TFIID

#### Modules in Preinitiation Complex

Cyril Esnault, Yad Ghavi-Helm, Sylvain Brun, Julie Soutourina, Nynke Van Berkum, Claire Boschiero, Frank Holstege, and Michel Werner

### Supplemental Experimental Procedures

#### Plasmid constructions and cloning

All cloning were done using the Gateway Invitrogen cloning method. Wild-type *MED11* gene was amplified from YPH499 genomic DNA using oligonucleotides matching the gene sequence initiation codon and following codons for the 5' forward primer and the stop codon and preceding codons for the 3' reverse primer. The oligonucleotides were flanked with *attB1* or *attB2* sequences, respectively. The amplified sequence was cloned into pDONR201 (Invitrogen) using standard BP reaction. The recombinant plasmid was sequence verified. The cloned sequence was then transferred into pVV208 (*CEN URA3 pTetO7*) or pVV204 (*CEN TRP1 pTetO7*) vector (Van Mullem et al., 2003) by the LR reaction. *RAD3* was cloned into pVV221 (*2 $\mu$  URA3 pTetO7*) and the *MED11* alleles into pVV220 (*2 $\mu$  TRP1 pTetO7*), pVV212 (derived from pGBT9) and pVV213 (derived from pACTII) using a similar procedure. K7N, V68D and G108S point mutations in Med11 were obtained by PCR overlap extension mutagenesis (Higuchi et al., 1988), cloned using the Gateway standard method and transferred into pVV204.

#### Yeast strains

Strains are listed in Supplementary Table 2. *MED11* was deleted in YPH499 (*MATa ura3-52 his3-200 ade2-101 trp1-63 lys2-801 leu2-1*) complemented with pVV208-*MED11* and replaced by a *Kan-MX6* marker using the standard one-step method (Longtine et al., 1998).

For ChIP experiments, the Med5-3HA, Tfa1-3HA, Ccl1-3HA, Ssl1-3HA, Rad3-3HA and Kin28-3HA strains were derived from yCE001 (*MATa ura3-52 his3-200 ade2-101 trp1-63 lys2-801 leu2-1 med11:Kan-MX6/CEN URA3 pTetO<sub>7</sub>::MED11*). The tagged strains were obtained by insertion of 3HA epitopes followed by a *his3MX6* marker using the method of Longtine et al (Longtine et al., 1998). The sequence was inserted between the last codon and the stop codon of the target ORF. The yCE001-8 strains were transformed with the pVV204-*MED11*, pVV204-*med11-T47A*, pVV204-*med11-L82P* and pVV204-*med11-G108S* plasmids. The pVV208-*MED11* was then chased on 5-FOA medium.

### **Random mutagenesis**

*MED11* was amplified from pVV208-*MED11* by PCR with pVV208-forward and pVV208-reverse oligos. After 10 cycles of PCR, 1mM MnCl<sub>2</sub> was added and 30 PCR cycles were performed. The PCR products were transformed with pVV204 linearized by *BspE1* in yCE001. The pVV208-*MED11* plasmid was counter-selected on 5-FOA medium. Colonies of the mutagenized strain were then replicated on YPD at 30°C or at 37°C for three days and thermosensitive mutants were selected. All the plasmids were isolated, re-transformed in yCE001 and sequenced.

### **Yeast two-hybrid assays**

The two-hybrid assays were performed in Y190 (*MATa gal4 gal80 his3 trp1-901 ade2-101 ura3-52 leu2-3,112 URA3::GAL1::lacZ LYS2::GAL4(UAS)::HIS3 cyhR*). The  $\beta$ -galactosidase activity was revealed by an X-Gal overlay assay (Werner et al., 1993).  $\beta$ -galactosidase was assayed on three clones for each experiment by the classical method of Miller (Miller, 1972).

### **Coimmunoprecipitations**

Whole-cell extracts were prepared from 100 ml of cells growing exponentially in yeast YPD medium. Cells were collected at an optical density of 0.6 to 0.8 at 600 nm ( $OD_{600}$ ), washed twice with water and twice with lysis buffer (50 mM HEPES [pH 7.5], 100 to 400 mM NaCl, 20% glycerol, 1 mM dithiothreitol [DTT], 0.05% NP-40) supplemented with a protease inhibitor cocktail (Complete; Roche) and 1 mM phenylmethylsulfonyl fluoride, and resuspended in 0.5 ml of the same buffer. Lysis was performed in the presence of glass beads (0.2 ml, 425 to 600  $\mu$ m) by vortexing for 30 min at 4°C. The cell debris were eliminated by centrifugation (15 min at 4°C at 18,000 x g, twice). The supernatant was collected and stored at -80°C.

During the immunoprecipitation procedure, the incubations and washes were performed at 4°C with agitation. A total of  $2 \times 10^7$  anti-mouse immunoglobulin G (IgG)-magnetic beads (DynaM450) were washed with 0.1% bovine serum albumin in phosphate-buffered saline (PBS; 137 mM NaCl, 2.7 mM KCl, 10 mM  $Na_2HPO_4$ , 2 mM  $KH_2PO_4$ ) and preincubated for 30 min in 0.1% bovine serum albumin in PBS. We used the beads with 8WG16 antibodies for Pol II, with 12CA5 for Med5-3HA or alone for TAP immunoprecipitation. The 8WG16 antibodies (200 ng/ $\mu$ l) were added for 1 h, after which the beads were washed three times for 5 min and twice shortly with lysis buffer. The protein extracts (15  $\mu$ g/ $\mu$ l, 100  $\mu$ l) were incubated with the beads for 3 h,

after which the beads were washed three times (5 min) with lysis buffer. The affinity-purified proteins were released from the beads by boiling for 10 min. Eluted proteins were separated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and analyzed by Western blotting with 12CA5 anti-HA, PAP Peroxydase anti-Peroxydase, 8WG16 anti-CTD of Rpb1 (Covance), anti-Med6, anti-Med17, anti-Med-20 (gift from S. Hahn), anti-Med8 (gift from S. Björklund) or anti-Med18 and anti-Med14 (gift from R. Kornberg).

### **Chromatin immunoprecipitations**

100 ml cell cultures were grown exponentially to  $OD_{600}=0.6$ , shifted for 45 minutes to 37°C and cross-linked with 1 % formaldehyde for 10 min. The 3HA-tagged proteins were immunoprecipitated with 12CA5 antibody bound to IgG magnetic beads (Dynabead). Pol II was immunoprecipitated with 8WG16 anti-CTD antibody (Covance). Rabbit antibody to recombinant yeast TBP was a gift from A. Weil. Immune complexes were washed as described previously (Kuras and Struhl, 1999). Elution of immunoprecipitated chromatin, crosslink reversal and DNA purification were performed as described (Kuras et al., 2003). Immunoprecipitated DNA was analysed by quantitative real time PCR on an ABI Prism 7000 or 7300 machine (Applied Biosystem). The PCR reactions were carried out in 25 µl containing 0.4 µM of each primer, and 12.5 µL of Mastermix SYBR green PCR reaction (Applied Biosystem). Relative quantification using a standard curve method was performed and the occupancy level for a specific fragment was defined as the ratio of immunoprecipitated DNA over total DNA. In each case, three independent experiments were averaged. The corresponding standard deviations are indicated in

the Figures. The cells were cultivated in YPD medium, *GAL1* ORF region was used as a non-transcribed control.

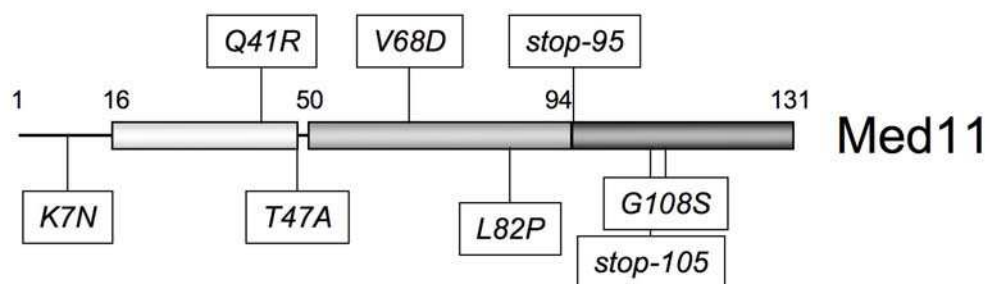
### **Kinase assay**

TFIIH was immunoprecipitated with 12CA5 anti-HA antibody as described above from yCE002 strain. Strains deriving from yCE006 expressing Med5-3HA in wild-type or *med11-T47A* mutant background were grown for 45 min at 37°C in YPD. Mediator was immunoprecipitated from the extracts with 12CA5 antibody. Immunoprecipitated complexes, TFIIH or Mediator, were eluted in 50 µl of kinase buffer (20 mM Tris pH 7.5, 50 mM KCl, 5 mM MgCl<sub>2</sub>, 2.5 mM MnCl<sub>2</sub>, 1 mM dithiotreitol (DTT)) with 0.5 µg/µl of 3HA peptide. Kinase reactions were then performed with 30 µl of each purified complex supplemented with 10 µCi of [ $\alpha$ -<sup>32</sup>P] ATP and 4 µg of CTD peptide for 30 min at 30°C as previously described (Hermand et al., 1998). As a negative control for CTD phosphorylation stimulation by Mediator, we used immunoprecipitated extracts from mutant strains in which Med5 was untagged. To control for non-specific phosphorylation by TFIIH, we performed the immunoprecipitation procedure from a strain in which TFIIH was not tagged and used the proteins adsorbed non-specifically instead of immunoprecipitated TFIIH.

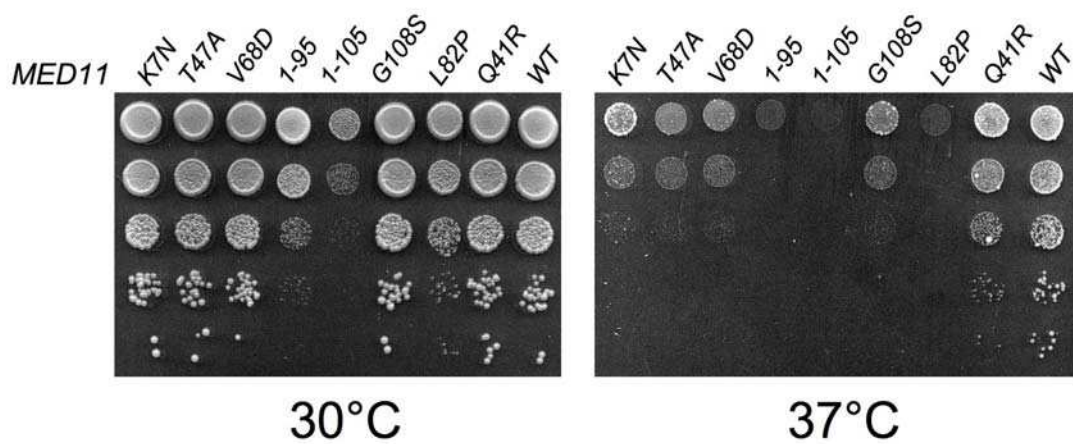


Figure S1

A



B



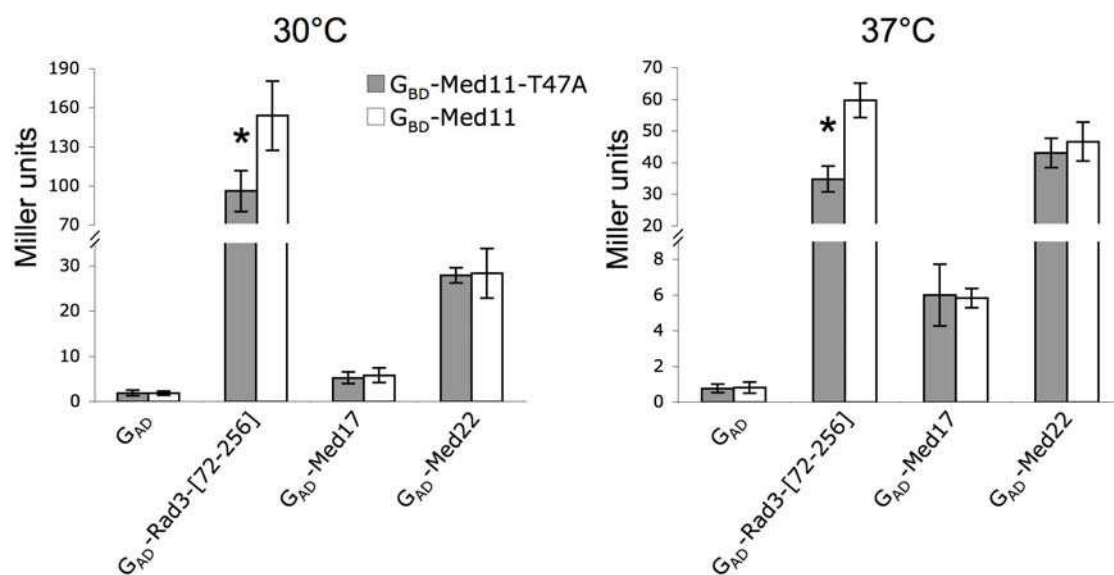
**Figure S1. *med11* thermosensitive mutants.**

(A) Schematic representation of the localization of Med11 point mutations. The grey boxes indicate the positions of the conserved domains.

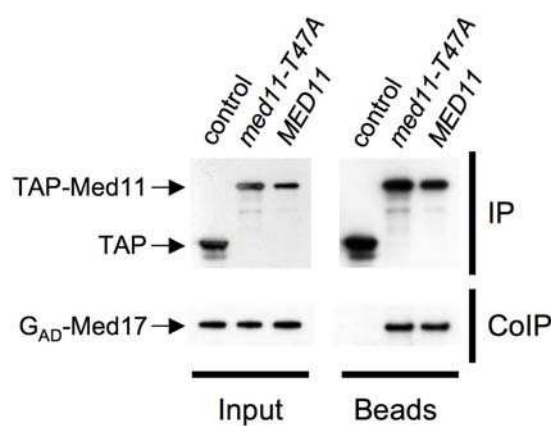
(B) Thermosensitive phenotype of the *med11* mutants. Cultures of wild-type (WT) and mutant *med11* yeast strains were serially diluted and spotted on YPD agar plates. Each plate was incubated for 3 days at either the permissive (30°C) or non-permissive (37°C) temperature.

Figure S2

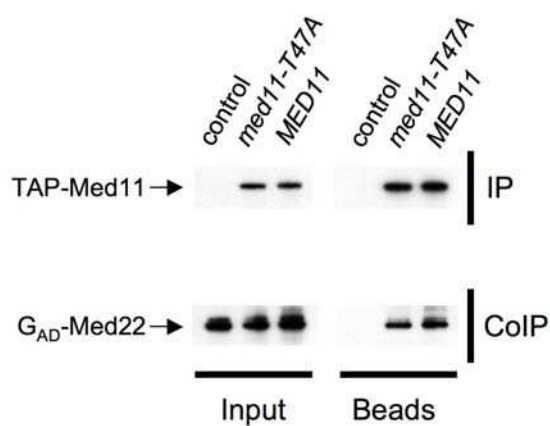
A



B



C



**Figure S2. *med11-T47A* mutation affects specifically the interaction between Med11 and Rad3-[72-256].**

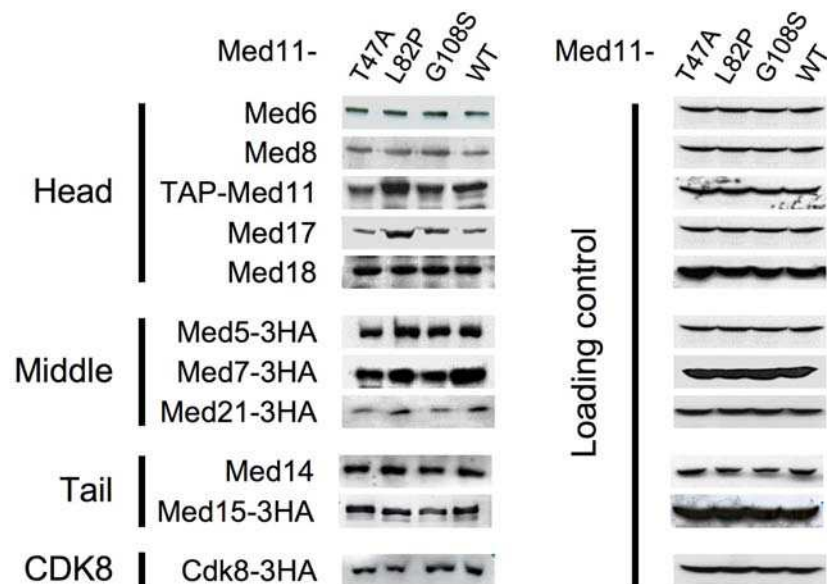
(A) *med11-T47A* affects specifically the interaction between Med11 and Rad3 in two-hybrids assays. Two-hybrid assays were performed in Y190 strains transformed by pGBT9-Med11 or pGBT9-Med11-T47A ( $G_{BD}$ -Med11 or  $G_{BD}$ -Med11-T47A), and pACTII, pACTII-Med17, pACTII-Med22 or pACTII-Rad3-[72-256] ( $G_{AD}$ ,  $G_{AD}$ -Med17,  $G_{AD}$ -Med22 or  $G_{AD}$ -Rad3-[72-256]).  $\beta$ -galactosidase was assayed according to the Miller method from extract of strains grown at 30°C or at 37°C (Miller, 1972). Values represent the average of three independent experiments. Error bars indicate standard deviation. \* represents a significant difference between the wild-type and the mutant at p-value <0.05 in a Student t-test.

(B) *med11-T47A* does not affect the interaction between Med11 and Med17 in CoIP assays. Strains expressing TAP, TAP-Med11 or TAP-Med11-T47A together with  $G_{AD}$ -Med17 (the  $G_{AD}$  contains an HA tag) were grown to exponential phase at 30°C. Protein extracts of these strains were immunoprecipitated in a buffer containing 150 mM NaCl with magnetic beads coupled to mouse IgG (Invitrogen). After immunoprecipitation, the proteins were revealed by Western blotting using 12CA5 or PAP antibodies.

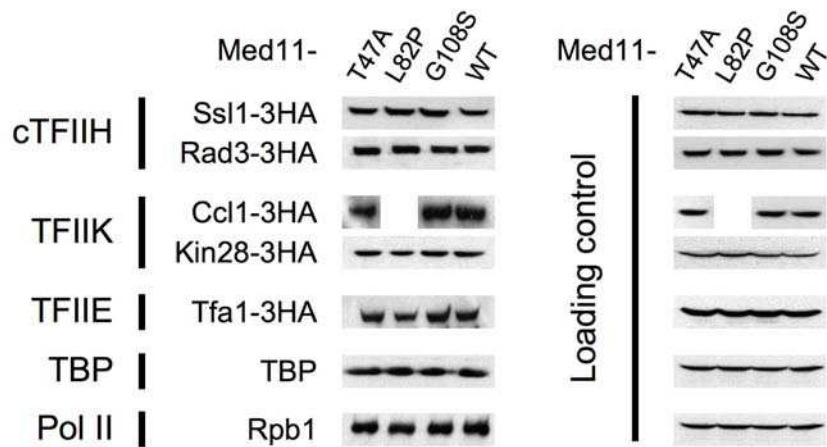
(C) *med11-T47A* does not affect the interaction between Med11 and Med22 in CoIP assays. Strains expressing  $G_{AD}$ -Med22 alone or together with TAP-Med11 or TAP-Med11-T47A were grown to exponential phase at 30°C. Protein extracts of these strains were immunoprecipitated in a buffer containing 150 mM NaCl with magnetic beads coupled to mouse IgG (Invitrogen). After immunoprecipitation, the proteins were revealed by Western blotting using 12CA5 or PAP antibodies.

# Figure S3

## A



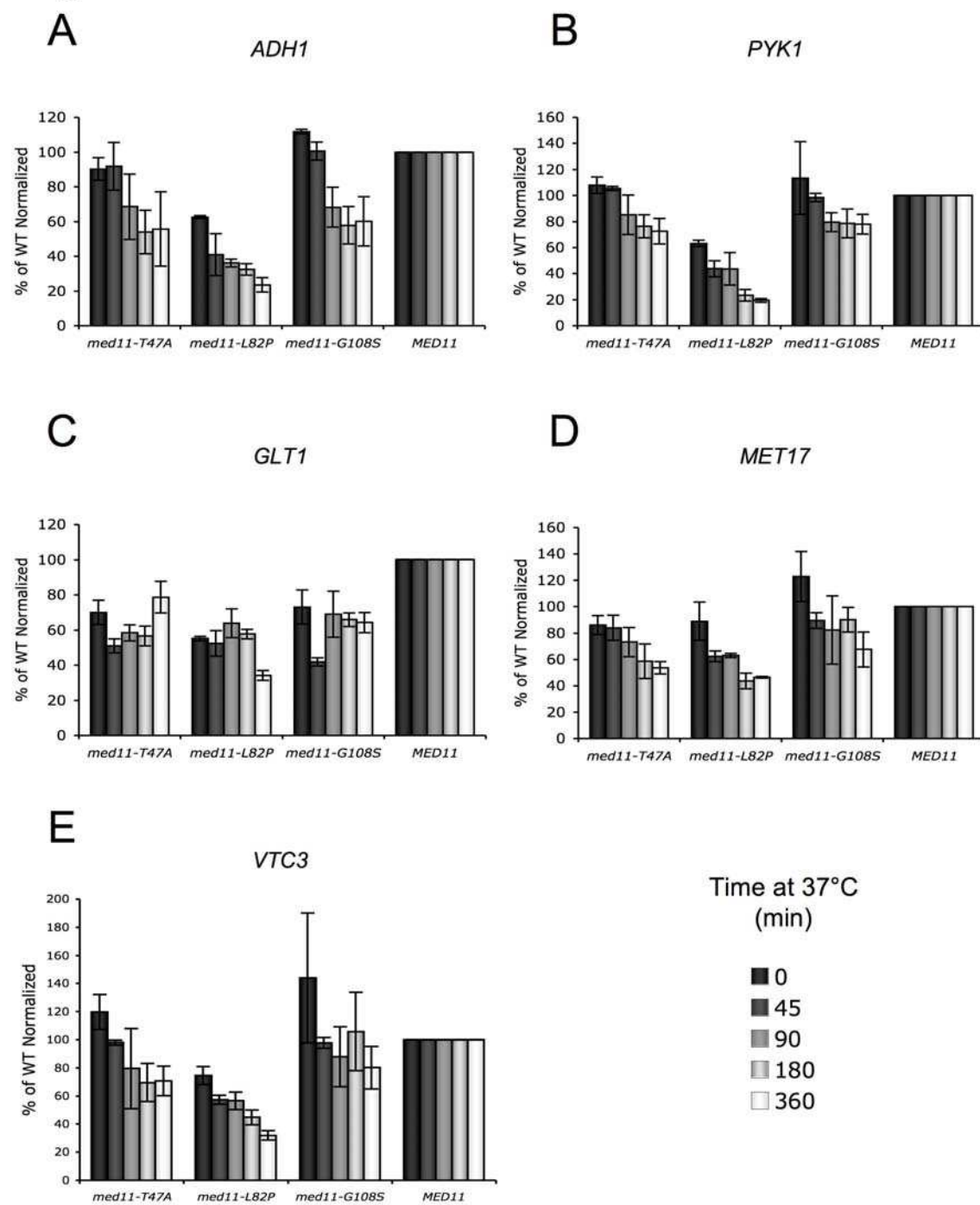
## B



**Figure S3. *med11* mutations do not affect the abundance of subunits of the transcription machinery.**

Yeast strains were grown for 45 minutes at 37°C. Protein extracts from strains expressing TAP-Med11, Med5-3HA, Med7-3HA, Med21-3HA, Med15-3HA, Cdk8-3HA, , Ssl1-3HA, Rad3-3HA, Ccl1-3HA, Kin28-3HA or Tfa1-3HA were separated by SDS-PAGE and analyzed by Western blotting using anti-HA (12CA5) or Peroxydase anti-Peroxydase (PAP) antibodies. Med6, Med8, Med17, Med18, Med14, TBP and Rpb1 levels were assessed using anti-Med6, anti-Med8, antiMed17, anti-Med18, anti-Med14, anti-TBP or anti-Rpb1 antibodies, respectively, in cell extracts from the Med5-3HA strain. Western blots with an antibody against the vacuolar ATPase were performed to control gel loading. (A) Western blots of Mediator subunits. (B) Western blots of GTFs and Pol II subunits.

Figure S4

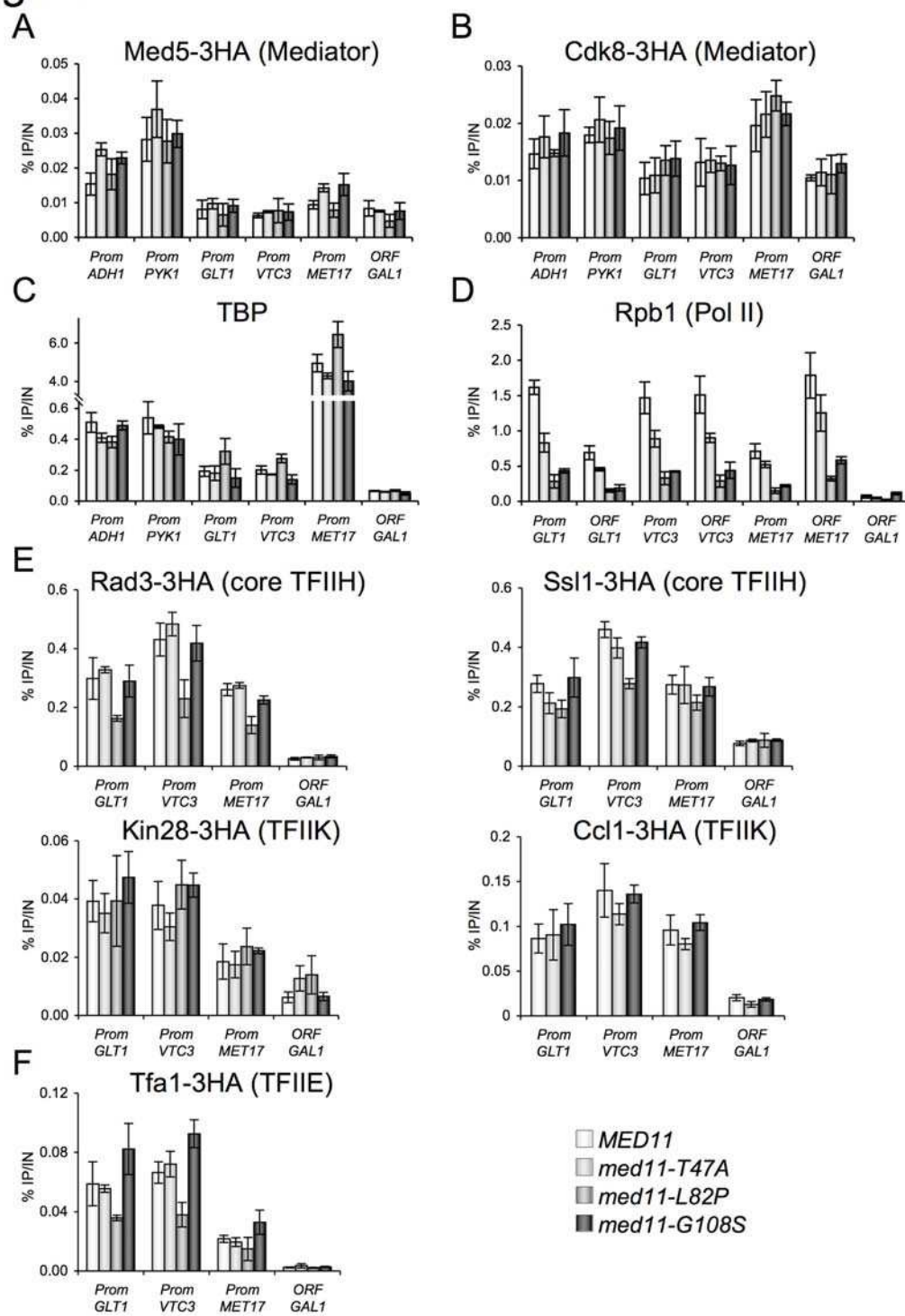


**Figure S4. *MED11* alleles influence mRNA abundance after a temperature shift at 37°C.**

Yeast strains expressing Med11, Med11-T47A, Med11-L82P or Med11-G108S were grown at 30°C and then shifted for the indicated time at 37°C in YPD medium. mRNA levels were determined by quantitative RT-PCR. RNA was extracted as described by Schmitt et al. (Schmitt et al., 1990). Reverse transcription of 5 µg RNA samples was performed in the presence of Super Script II and random primers (Invitrogen) for 1-2 h at 42°C in appropriate buffer. Q-PCR results were normalized using 25S rRNA and to the wild-type signal. Values represent the average of two independent experiments, and error bars indicate standard deviations. 25S rRNA was used as internal control for normalization.



Figure S5



**Figure S5. Effect of *med11* mutations on Mediator, TBP, Pol II, TFIIE, cTFIIH and TFIIK occupancy.**

Standard ChIP assays were performed on chromatin extracts as previously described in Figure 3 using anti-HA (12CA5), anti TBP or anti Rpb1 CTD (8WG16) antibodies except that the protein occupancies were examined on *ADH1*, *PYK1*, *GLT1*, *VTC3* and *MET17* (A-C) or *GLT1*, *VTC3* and *MET17* only (D-F). Immunoprecipitated DNA were amplified with primers as indicated in supplementary Table 1. Values represent the average of three independent experiments. Error bars indicate the standard deviation.

(A) Mediator occupancy. Strains expressing Med5-3HA in wild-type or mutant backgrounds were grown for 45 minutes at 37°C.

(B) CDK module occupancy. Strains expressing Cdk8-3HA in wild-type or mutant backgrounds were grown for 45 minutes at 37°C.

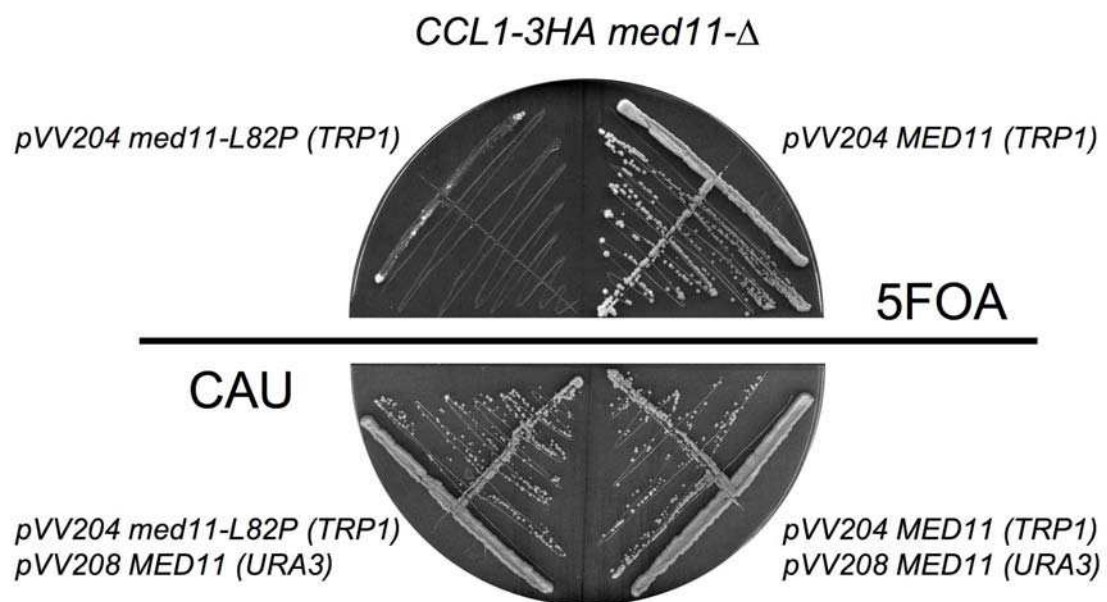
(C) TBP occupancy. Wild-type or mutant strains were grown for 45 minutes at 37°C.

(D) Pol II occupancy. Wild-type or mutant strains were grown for 45 minutes at 37°C as in Figure 3.

(E) cTFIIH and TFIIK subunits occupancy. Strains expressing Rad3-3HA, Ssl1-3HA, Kin28-3HA or Ccl1-3HA in wild-type or mutant backgrounds were grown for 45 minutes at 37°C.

(F) TFIIE occupancy. Strains expressing Tfa1-3HA in wild-type or mutant backgrounds were grown for 45 minutes at 37°C.

Figure S6

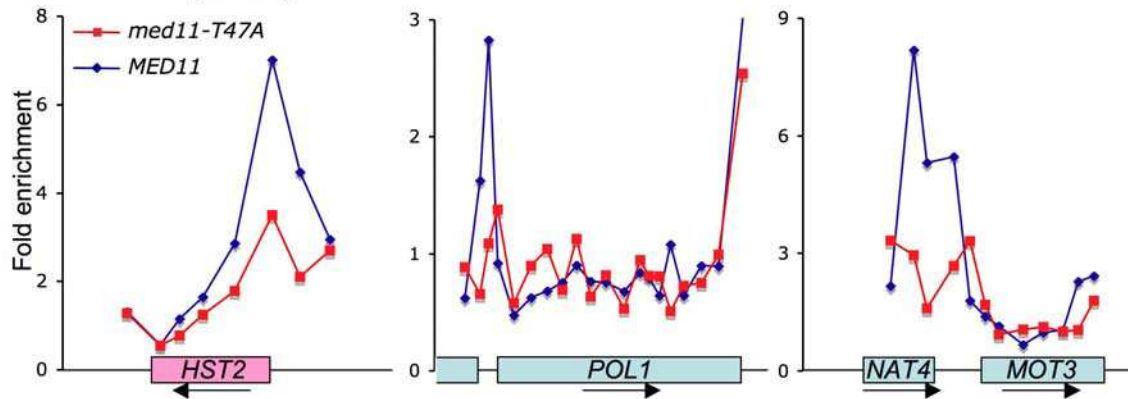


**Figure S6. *med11-L82P* and *cc11-3HA* are co-lethal.**

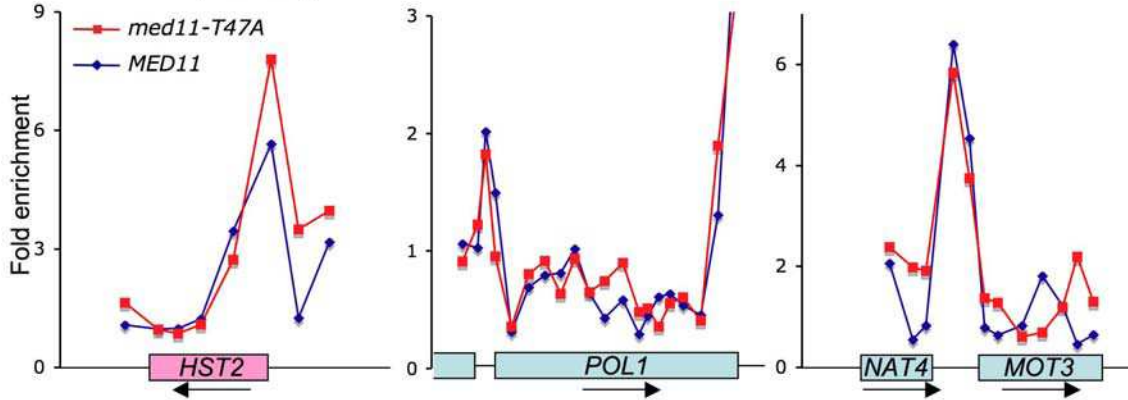
Yeast strain yCE007 with the indicated genotype carrying plasmids expressing *MED11* or *med11-L82P* were grown on selective medium CAU or on medium containing 5-FOA to select against the wild-type *MED11* plasmid (pVV208; *CEN*, *URA3*). The plates were incubated at 30° for three days.

## Figure S7

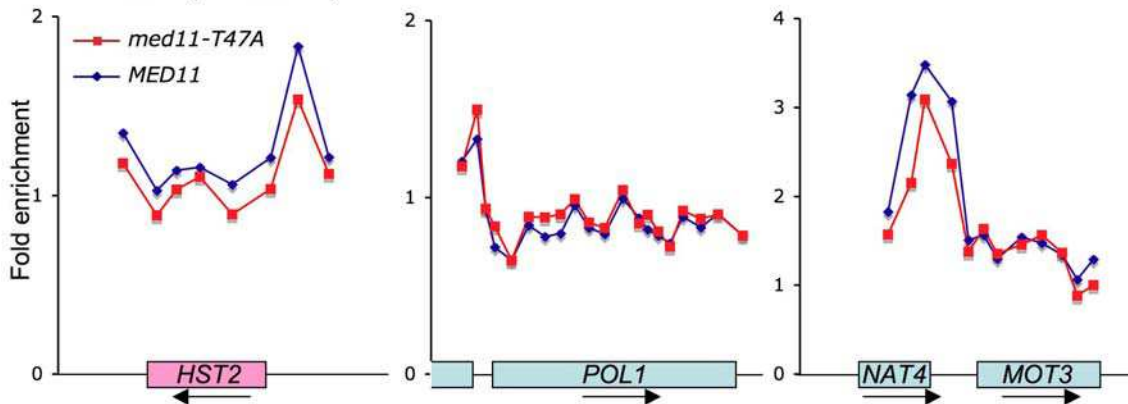
Kin28-3HA (TFIIK)



RAD3-3HA (cTFIIH)



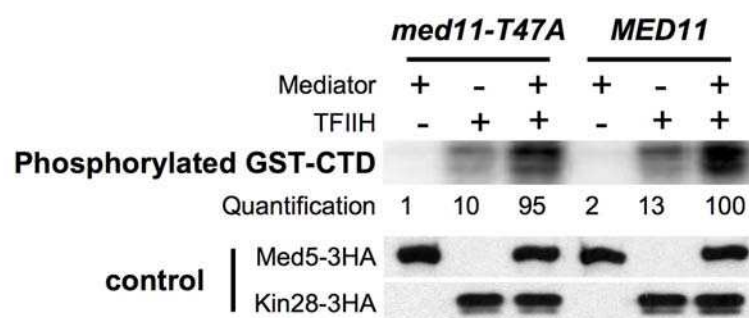
Med5-3HA (Mediator)



**Figure S7. Examples of occupancy profile changes of Kin28 in *med11-T47A* mutant.**

Examples of genomic regions showing reduced TFIIK binding. ChIP-chip assays were performed using antibodies against the HA tag (12CA5). The graphics present binding ratio of Kin28-3HA, Rad3-3HA and Med5-3HA.

Figure S8



**Figure S8. Effect of *med11-T47A* mutation on stimulation of CTD phosphorylation by Mediator *in vitro*.**

Strains expressing Med5-3HA in wild-type or *med11-T47A* background were grown in YPD rich media for 45 minutes at 37°C. Mediator was prepared by immunoprecipitation with anti-HA antibody (12CA5). As a negative control, we applied the immunoprecipitation protocol to untagged wild-type or mutant strains. TFIIF was prepared by immunoprecipitation from a Kin28-3HA strain. An untagged strain was used as a negative control. The phosphorylation assays were performed with immunoprecipitated TFIIF, wild-type or mutant immunoprecipitated Mediator, 4 µg recombinant GST-CTD and [ $\alpha$ -<sup>32</sup>P] ATP for 30 min at 30°C. In the bottom panel, protein level in the assay was verified by Western blotting using anti-HA antibody.



Table S1. Oligonucleotides

	Name	Forward	Reverse
Cloning	<i>MED11</i>	GGGGACAAGTTTGTACAAAAAAG CAGGCTTGATGCAGGTTCTCAAT ACGAAG	GGGGACCACTTTGTACAAGAAA GCTGGGTCTTACTTGGATTTCG AAGGATC
	<i>RAD3</i>	GGGGACAAGTTTGTACAAAAAAG CAGGCTTGATGAAGTTTATATAG ATGATTACC	GGGGACCACTTTGTACAAGAAA GCTGGGTCTCACTGCATTCTA TATCTTCATCTTC
	<i>rad3-[72-256]</i>	GGGGACAAGTTTGTACAAAAAAG CAGGCTTGATGTCTCGTACTATG TCTGAAATTG	GGGGACCACTTTGTACAAGAAA GCTGGGTCCAATGCATCCGTTG TCAAGTC
	<i>RAD3-no_stop</i>	GGGGACAAGTTTGTACAAAAAAG CAGGCTTGATGAAGTTTATATAG ATGATTACC	GGGGACCACTTTGTACAAGAAA GCTGGGTCTGCATTCTATAT CTTCATCTTC
Mutagenesis	PVV208	ACTAAATTACCGGATCAATTCTG	CTAATTACATGATGCGGCCCC
ChIP and RT-PCR	PROM <i>ADH1</i>	ATAGGCGCATGCAACTTCTT	CATCAGCTCTGGAACAACGA
	<i>ADH1-1</i>		
	PROM <i>ADH1</i>	TTCCTTCATTACGCACACT	AGGGAACGAGAACAATGACG
	<i>ADH1-2</i>		
	ORF <i>ADH1</i>	GGGTATTGACGGTGGTGAAG	AAACGTTGATGACACCGTGA
	<i>ADH1-3</i>		
	ORF <i>ADH1</i>	GGCTGGAAGATCGGTGACTA	TCAGCGGTAGCGTATTGTTG
	<i>ADH1-4</i>		
	ORF <i>ADH1</i>	GGTGCCAAGTGTTGTTCTGA	GACAAGCCGACAACCTTGAT
	<i>ADH1-5</i>		
	PROM <i>PYK1</i>	CGCACCGTCACAAAGTGTT	TGGGAAGGAAAGGAAATCAC
	<i>PYK1-1</i>		
	PROM <i>PYK1</i>	CCTTTCCTTCCCATATGATGC	ACTTTGAAAGGGGACCATGA
	<i>PYK1-2</i>		
	ORF <i>PYK1</i>	TGCTTTGAGAAAGGCTGGTT	TCTGATTTCTGGACCCTTGG
	<i>PYK1-3</i>		
	ORF <i>PYK1</i>	ATGGTTGCCAGAGGTGACTT	TCTGGTTGGTCTTGGGTTGT
	<i>PYK1-4</i>		
	ORF <i>PYK1</i>	CAGAGGTGTCTTCCCATTCG	CCTTGAAACCTTGGATGGAA
	<i>PYK1-5</i>		
	PROM <i>GLT1</i>	GAAGCAATAAATTGCCCAAAA	TGGAGACGATTAGTTTGGTGTG
	ORF <i>GLT1</i>	GCGTTGGTCCATTCAAGATT	GAGCGCATCCAATTCTTGTT
	PROM <i>VTC3</i>	TCGGCGGACTCCAATACTATTA	ACAGGGA AAAATGTGCCAAG
	ORF <i>VTC3</i>	CCTCCCTGGAAGGATTCATA	CTTTATCCAAGGCCTCCACA
	PROM <i>MET17</i>	GGGTTCGAATCCCTTAGCTC	GATGGCGCCTTGACTAACTT
	ORF <i>MET17</i>	TACCCCCATCCATACAATGC	GAAGTGGTGGCGTAAATTGG
	ORF <i>GAL1</i>	AAAGAAACTTGCACCGGAAA	GGCCCATATTCGCTTTAACA
	PROM <i>HST2</i>	CTTTCTGCCGTCCAGCTATG	CTCAGTGACGGGCACAGATA
	<i>HST2-1</i>		
	ORF <i>HST2</i>	CGATCTCTCCGTGTTTCGAT	CGGAGATCGGAAGTCTGGTA
	<i>HST2-2</i>		
	ORF <i>HST2</i>	TGGCAGTTTTGCTCACTGTC	AACTATTGCCGGCTTTACCA
	<i>HST2-3</i>		

ORF <i>HST2</i>	CGTTTATCCCTTTGCGTCTC	AGTTGCTCGGCAAACATCATC
<i>HST2-4</i>		
ORF <i>HST2</i>	GAAGCTACAGCGTCTCAATGG	CGCGCATGCTAGATGAATAA
<i>HST2-5</i>		
25S <i>rRNA</i>	GGTTATATGCCGCCCCGTCTTGA	CCCAACAGCTATGCTCTTACTC

---

Table S2. Yeast strains

Name	Tagged protein	Genotype	Reference
Y190		<i>MATa gal4 gal80 his3 trp1-901 ade2-101 ura3-52 leu2-3,112 URA3::GAL1::lacZ LYS2::GAL4(UAS)::HIS3 cyhRY</i>	(Harper et al., 1993)
YPH499		<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1</i>	(Sikorski and Hieter, 1989)
YPH500		<i>MATα ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1</i>	(Sikorski and Hieter, 1989)
YCE001		<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 Δmed11::kan / CEN URA3 pTetO7::MED11</i>	This work
YCE002	Med5-3HA	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 Δmed11::kan MED5:3HA:HIS3MX6 / CEN URA3 pTetO7::MED11</i>	This work
YCE003	Med15-3HA	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 Δmed11::kan MED15:3HA:HIS3MX6 / CEN URA3 pTetO7::MED11</i>	This work
YCE004	Rad3-3HA	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 Δmed11::kan RAD3:3HA:HIS3MX6 / CEN URA3 pTetO7::MED11</i>	This work
YCE005	Ssl1-3HA	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 Δmed11::kan SSL1:3HA:HIS3MX6 / CEN URA3 pTetO7::MED11</i>	This work
YCE006	Kin28-3HA	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 Δmed11::kan KIN28:3HA:HIS3MX6 / CEN URA3 pTetO7::MED11</i>	This work
YCE007	Ccl1-3HA	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 Δmed11::kan CCL1:3HA:HIS3MX6 / CEN URA3 pTetO7::MED11</i>	This work
YCE008	Tfa1-3HA	<i>MATa ura3-52 his3-Δ200 ade2-101uaa trp1-Δ63 lys2-801uag leu2-Δ1 Δmed11::kan TFA1:3HA:HIS3MX6 / CEN URA3 pTetO7::MED11</i>	This work

## Supplemental References

Harper, J. W., Adami, G. R., Wei, N., Keyomarsi, K., and Elledge, S. J. (1993). The p21 Cdk-interacting protein Cip1 is a potent inhibitor of G1 cyclin-dependent kinases. *Cell* 75, 805-816.

Hermant, D., Pihlak, A., Westerling, T., Damagnez, V., Vandenhaute, J., Cottarel, G., and Makela, T. P. (1998). Fission yeast Csk1 is a CAK-activating kinase (CAKAK). *Embo J* 17, 7230-7238.

Higuchi, R., Krummel, B., and Saiki, R. K. (1988). A general method of in vitro preparation and specific mutagenesis of DNA fragments: study of protein and DNA interactions. *Nucleic Acids Res* 16, 7351-7367.

Kuras, L., Borggrefe, T., and Kornberg, R. D. (2003). Association of the Mediator complex with enhancers of active genes. *Proc Natl Acad Sci U S A* 100, 13887-13891.

Kuras, L., and Struhl, K. (1999). Binding of TBP to promoters in vivo is stimulated by activators and requires Pol II holoenzyme. *Nature* 399, 609-613.

Longtine, M. S., McKenzie, A., III, Demarini, D. J., Shah, N. G., Wach, A., Brachat, A., Philippsen, P., and Pringle, J. R. (1998). Additional modules for versatile and economical PCR-based gene deletion and modification in *Saccharomyces cerevisiae*. *Yeast* 14, 953-961.

Miller, J. H. (1972). In *Molecular Genetics* (Cold Spring Harbor, NY, Cold Spring Harbor Laboratory Press), pp. 352-355.

Schmitt, M. E., Brown, T. A., and Trumpower, B. L. (1990). A rapid and simple method for preparation of RNA from *Saccharomyces cerevisiae*. *Nucleic Acids Res* 18, 3091-3092.

Sikorski, R. S., and Hieter, P. (1989). A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. *Genetics* 122, 19-27.

Van Mullem, V., Wery, M., De Bolle, X., and Vandenhaute, J. (2003). Construction of a set of *Saccharomyces cerevisiae* vectors designed for recombinational cloning. *Yeast* 20, 739-746.

Werner, M., Chaussivert, N., Willis, I. M., and Sentenac, A. (1993). Interaction between a complex of RNA polymerase III subunits and the 70 kDa component of TFIIIB. *J Biol Chem* 268.



### ***III/ L'ARN polymérase II: senseur de la disponibilité en nucléosides triphosphates***

#### Contexte du travail

Ce travail a été récemment publié dans EMBO J. (Kwapisz et al., 2008). Il est largement basé sur une approche génétique initiée par François Lacroute, et poursuivie par Marta Kwapisz et Max Wéry, qui ont en outre réalisé les expériences de CHIP local (*URA2*, *URA8* et *IMD2*) et de transcriptome. En ce qui me concerne, j'ai été plus particulièrement impliquée dans la réalisation et l'analyse des données de localisation à grande échelle de l'ARN polymérase II dans une souche sauvage et dans le mutant *rpb1-L1397S*.

La transcription du gène *URA2*, codant une enzyme clé de la voie de biosynthèse de l'UTP, intéresse particulièrement le laboratoire, du fait de son mode de régulation. En effet, sa transcription est régulée par la disponibilité en UTP; la transcription du gène est réprimée en présence d'UTP, et dé-réprimée en présence de 6-azauracile (6AU, inhibiteur de la biosynthèse d'UTP et de GTP) dans le milieu de culture. Etrangement, ce phénomène ne dépend pas de Ppr1, un activateur de la voie de biosynthèse des pyrimidines responsable de la régulation transcriptionnelle des gènes *URA1* ou *URA3*.

La recherche de mutants capables de mimer une carence en UTP a, dans un premier temps, été réalisée grâce à la construction de deux vecteurs comportant les gènes rapporteurs *LacZ* ou *HIS3* sous le contrôle du promoteur du gène *URA2* (défini ici comme la région intergénique en amont du gène *URA2*). L'analyse de l'organisation de la région promotrice du gène *URA2* révèle la présence d'un intron en amont de la région codante ainsi que la position du site d'initiation de la transcription. Une étude réalisée en parallèle montre la présence d'un deuxième site d'initiation de la transcription, à partir duquel est produit un court transcrit cryptique instable (CUT) dans la même orientation que le gène *URA2* impliqué dans la régulation négative de la transcription de ce gène (Thiebaut et al., 2008).

La construction de mutants comportant des délétions partielles de la région promotrice du gène *URA2* et l'analyse de leur effet sur la transcription du gène rapporteur *LacZ* nous a permis d'identifier une région, appelée  $\Delta$ up2, dont la délétion provoque une expression constitutive même en présence d'uracile. Des expériences d'immunoprécipitation de la

chromatine révèlent que, dans une souche sauvage en condition de répression, l'occupation de l'ARN polymérase II présente un profil atypique, comparable à une situation de pause au promoteur. De plus, l'activation du gène *URA2* ne résulte pas en une augmentation du recrutement de l'ARN polymérase II au promoteur, mais en une augmentation de son occupation sur la région codante du gène. L'analyse de l'effet de la mutation  $\Delta$ up2 suggère que cette région bloque la progression de l'ARN polymérase II le long de l'ORF, résultant ainsi en une répression de la transcription du gène *URA2*.

Nous avons ensuite recherché par mutagenèse UV, des mutants exprimant le rapporteur *pURA::HIS3* de façon constitutive. De façon surprenante, les quatre mutations obtenues sont toutes situées au niveau d'une boucle conservée (Switch 1) de la sous-unité Rpb1 de l'ARN polymérase II. Nous nous sommes plus spécifiquement intéressés à la mutation *rpb1-L1397S*, ayant la particularité de présenter des effets épistatiques avec la délétion de la région  $\Delta$ up2. Ces deux mutants sont ainsi probablement affectés dans le même mécanisme régulant négativement l'expression du gène *URA2* en réponse à l'uracile.

Le mutant *rpb1-L1397S* présente un phénotype de croissance cellulaire à 30°C, ainsi qu'une diminution globale de la transcription et de la présence d'ARN polymérase II sur son génome par rapport à une souche sauvage. Cependant, l'analyse fine des résultats de ChIP-on-chip et de transcriptome obtenus chez ce mutant révèle, pour certains gènes tels *URA2*, une augmentation de l'occupation de l'ARN polymérase II sur l'ORF et du niveau d'ARNm correspondant par rapport à une souche sauvage. Ainsi, le mutant *rpb1-L1397S* active le gène *URA2* en permettant la présence de l'ARN polymérase II sur l'ORF, récapitulant ainsi l'effet de la délétion  $\Delta$ up2 ou de la présence de 6AU.

Ces expériences de localisation à grande échelle nous ont permis d'identifier plusieurs autres gènes, dont en particulier *IMD2* et *URA8*, des enzymes clé de la voie de biosynthèse du GTP et du CTP, respectivement. Il apparaît ainsi que l'ARN polymérase II, à travers la boucle Switch 1 de sa grande sous-unité Rpb1, pourrait être capable de sentir la présence de nucléosides triphosphates disponibles pour la transcription.

L'ensemble de ces résultats indique que la transcription du gène *URA2* est activée en absence d'UTP en permettant à l'ARN polymérase II de transcrire l'entièreté du gène *URA2*. En condition de répression, la transcription se limite essentiellement à la production d'un ARN cryptique dont le site de démarrage se trouve environ 100 paires de bases en amont de celui d'*URA2*. En condition d'activation (carence en UTP), l'ARN polymérase II continue de transcrire ce CUT (Thiebaut et al., 2008), mais devient en outre préférentiellement capable de

transcrire de façon processive le gène *URA2* à partir de son site initiateur. Ces deux modes de transcription dépendent de la même boîte TATA qui, de façon remarquable, est localisée environ 200 paires de bases en amont du site de démarrage de la transcription du gène *URA2* (Thiebaut et al., 2008), une observation difficile à concilier avec les modèles actuels de recrutement de l'ARN polymérase II au sein du PIC.

Le mécanisme transcriptionnel mis en jeu n'est pas réellement compris au stade actuel. Néanmoins, deux classes de mutants mimant une carence en UTP ont pu être identifiées: une délétion en *cis* du gène *URA2*, une mutation en *trans* dans le domaine Switch 1 de la sous-unité Rpb1 de l'ARN polymérase II. En plus d'activer la transcription du gène *URA2*, cette seconde mutation active également la transcription des gènes codant des enzymes clé de la voie de biosynthèse des autres nucléotides, suggérant que l'ARN polymérase II elle-même pourrait être le senseur de la disponibilité en nucléotides dans la cellule.

## Article





# Mutations of RNA polymerase II activate key genes of the nucleoside triphosphate biosynthetic pathways

This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits distribution, and reproduction in any medium, provided the original author and source are credited. This license does not permit commercial exploitation without specific permission.

Marta Kwapisz<sup>1,2,4</sup>, Maxime Wery<sup>1,4</sup>,  
Daphné Després<sup>1</sup>, Yad Ghavi-Helm<sup>1</sup>,  
Julie Soutourina<sup>1</sup>, Pierre Thuriaux<sup>1,\*</sup>  
and François Lacroute<sup>2,3</sup>

<sup>1</sup>CEA, iBiTec-S, Service de Biologie Intégrative et Génétique Moléculaire, Gif-sur-Yvette, France, <sup>2</sup>CNRS, Centre de Génétique Moléculaire, UPR2167, Gif-sur-Yvette, France and <sup>3</sup>Université Pierre et Marie Curie, Paris, France

The yeast *URA2* gene, encoding the rate-limiting enzyme of UTP biosynthesis, is transcriptionally activated by UTP shortage. In contrast to other genes of the UTP pathway, this activation is not governed by the Ppr1 activator. Moreover, it is not due to an increased recruitment of RNA polymerase II at the *URA2* promoter, but to its much more effective progression beyond the *URA2* mRNA start site(s). Regulatory mutants constitutively expressing *URA2* resulted from *cis*-acting deletions upstream of the transcription initiator region, or from amino-acid replacements altering the RNA polymerase II Switch 1 loop domain, such as *rpb1-L1397S*. These two mutation classes allowed RNA polymerase to progress downstream of the *URA2* mRNA start site(s). *rpb1-L1397S* had similar effects on *IMD2* (IMP dehydrogenase) and *URA8* (CTP synthase), and thus specifically activated the rate-limiting steps of UTP, GTP and CTP biosynthesis. These data suggest that the Switch 1 loop of RNA polymerase II, located at the downstream end of the transcription bubble, may operate as a specific sensor of the nucleoside triphosphates available for transcription.

The EMBO Journal (2008) 27, 2411–2421. doi:10.1038/emboj.2008.165; Published online 21 August 2008

Subject Categories: chromatin & transcription

Keywords: *IMD2*; *IMD3*; *S. cerevisiae*; *URA2*; *URA8*

steps of UTP biosynthesis are catalysed by Ura2, a bifunctional protein endowed with carbamoyl phosphate synthetase and aspartate transcarbamoylase activities (Potier *et al*, 1987). These are the main rate-limiting steps of the pathway, and their allosteric control by UTP is well documented (Lacroute *et al*, 1965; Serre *et al*, 2004). Similar to other genes of the UTP biosynthetic pathway, *URA2* transcription is induced in the presence of 6-azauracil, which depletes cells in UTP (Exinger and Lacroute, 1992). A previous study (Losson and Lacroute, 1981) has suggested that this does not depend on Ppr1, a transcriptional activator specific for the pyrimidine biosynthetic pathway (Loison *et al*, 1980; Losson and Lacroute, 1981; Flynn and Reece, 1999), as *ppr1-2*, a mutation unable to activate *URA1* or *URA3*, had no effect on Ura2 activity in cell-free extracts. Moreover, we here show that a *ppr1Δ* null allele remains fully competent for *URA2* transcription.

In this study, two classes of mutations were found to constitutively activate *pURA2::HIS3* or *pURA2::LacZ* reporter plasmids. The first class was due to short deletions immediately upstream of the *URA2* mRNA 5'-ends, indicating that this region acts as negative regulatory element of *URA2* transcription. A second class altered *RPB1*, which encodes the largest subunit of RNA polymerase II, and specifically modified the Switch 1 loop of the active site. One of these mutations, *rpb1-L1397S*, was investigated in more detail. It showed a genome-wide reduced occupancy of RNA polymerase II, consistent with its partial growth defect, but specifically activated *URA2*, *IMD2/IMD3* and *URA8* (encoding rate-limiting steps of the UTP, GTP and CTP biosynthetic pathways, respectively), thus raising the intriguing possibility that the Switch 1 loop of RNA polymerase II might act as a sensor of nucleoside triphosphate depletion.

## Results

### UTP depletion activates *URA2* independently of the Ppr1 activator

The *URA2* open reading frame is separated by 1231 nt from the stop codon of the upstream gene *TRK1* (Figure 1A). A *KpnI*–*Bam*HI cassette bearing this DNA was cloned in frame with the *LacZ* or *HIS3* open reading frames, generating the *pURA2::LacZ* reporter plasmids pFL80 and pFL80-H2, and the *pURA2::HIS3* reporter pFL81 (Materials and methods). There was a fivefold *LacZ* activation in wild-type cells grown under repressing (uracil) or derepressing (6-azauracil) conditions (Figure 1B), and a similar range of activation was observed in the steady-state level of *URA2* mRNAs (Figure 1C), indicating that the *pURA2::LacZ* reporter correctly reflects the transcriptional regulation of *URA2*.

## Introduction

DNA transcription depends on the availability of pyrimidine (UTP and CTP) and purine (GTP and ATP) ribonucleoside triphosphate substrates, and their *de novo* synthesis is subjected to tight homeostatic controls. In yeast, the first two

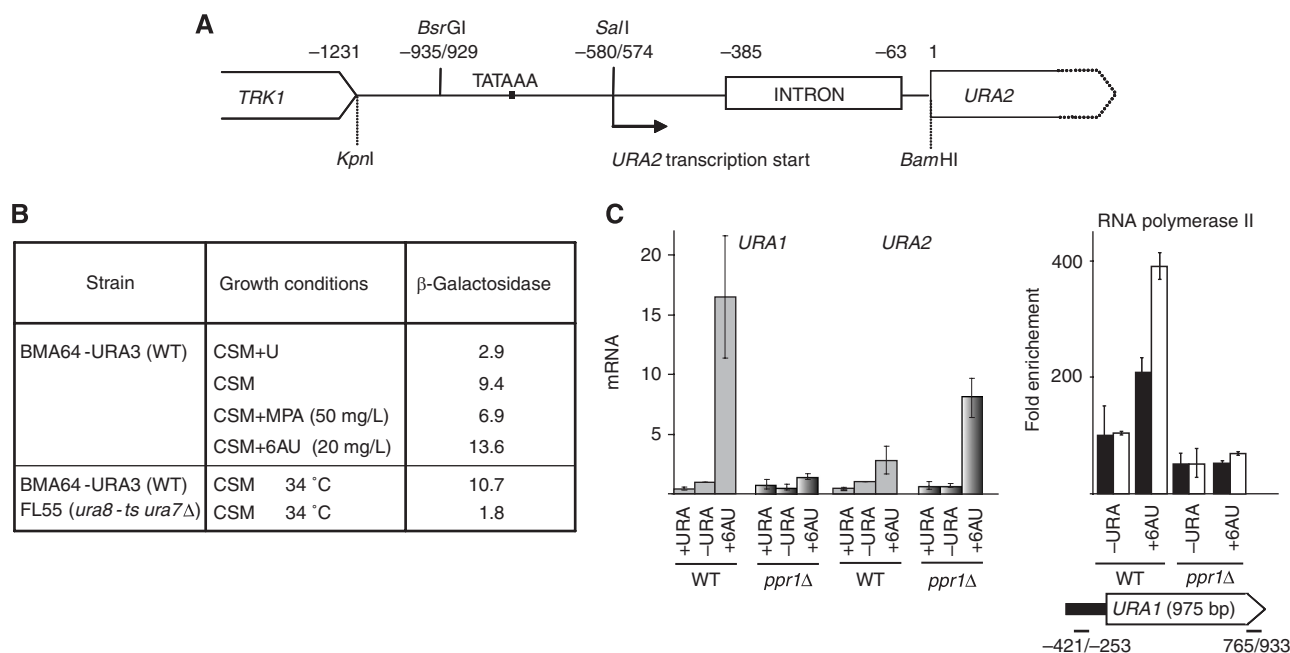
\*Corresponding author. CEA, iBiTec-S, Service de Biologie Intégrative et Génétique Moléculaire, Gif sur Yvette F-91191, France.

Tel.: +33 1 69 08 35 86; Fax: +33 1 69 08 47 12;

E-mail: pierre.thuriaux@cea.fr

<sup>4</sup>These authors contributed equally to this work

Received: 2 June 2008; accepted: 30 July 2008; published online: 21 August 2008



**Figure 1** Transcriptional regulation of *URA2* in response to UTP shortage. **(A)** Schematic view of the *TRK1*-*URA2* intergenic region. The corresponding DNA was cloned as a *KpnI*-*BamHI* cassette to generate the pFL80 and pFL80-H2 *pURA2::LacZ* reporter plasmids. A box denotes the 5'-UTR intron. The broken arrow corresponds to the *URA2* mRNAs 5'-ends. **(B)**  $\beta$ -galactosidase activities of BMA64-*URA3* (WT) and FL55 (*ura8-ts ura7Δ*) grown without uracil (CSM) at 30 and 34 °C, in the presence of uracil (CSM + U), mycophenolate (CSM + MPA) and 6-azauracil (CSM + 6AU). **(C)** Three independent cultures of BMA64-*URA3* (WT) and FL53 (*ppr1Δ*) were exponentially grown without (-URA) or with uracil (+URA), or with 200 mg/l of 6-azauracil (+6AU). *URA1* and *URA2* mRNAs were quantified by RT-PCR, and RNA polymerase II occupancy was determined by ChIP assays, using anti-CTD antibodies (8WG16). A schematic map represents the *URA1* gene, with the oligonucleotide pairs used as primers. Position 1 corresponds to the start codon.

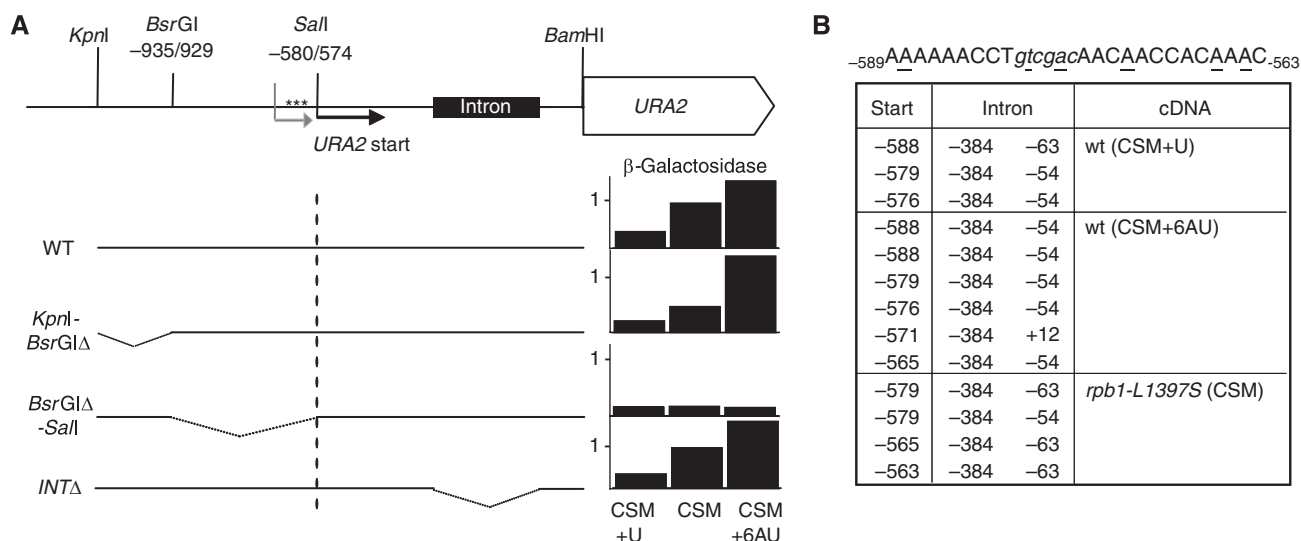
Adding 6-azauracil blocks UTP synthesis and also reduces the cellular pool of GTP (Exinger and Lacroute, 1992). However, no *pURA2::LacZ* activation was observed in the presence of mycophenolate, which specifically blocks IMP dehydrogenase (Shaw *et al*, 2001), and thus only depletes GTP. Likewise, counteracting GTP inhibition by adding guanine to the CSM+6AU medium did not reduce the activating effect of 6-azauracil (data not shown). Finally, impairing the metabolic conversion of UTP to CTP in temperature-sensitive mutant (*ura8-ts ura7Δ*) reduced the expression of *pURA2::LacZ*, which presumably reflects the accumulation of UTP under these conditions and implies that *URA2* does not respond to CTP shortage. These data indicated that *URA2* transcription is primarily activated by a shortage in UTP, with little or no effect of GTP or CTP.

Other genes of the pyrimidine pathway, such as *URA1* and *URA3*, are also derepressed under UTP depletion, and their activation is lost in *ppr1-2*, a mutation that alters the Ppr1 activator but has no effect on the Ura2 activity in cell-free extracts (Losson and Lacroute, 1981). Consistent with this observation, we found here that the *ppr1Δ* null allele strongly reduced the activation of *URA1* and *URA3* in the presence of 6-azauracil, as measured by RT-PCR quantification of their mRNAs and by their RNA polymerase II occupancy in chromatin-immunoprecipitation (ChIP) assays (shown in Figure 1C for the *URA1* gene). In contrast, *ppr1Δ* actually increased the level of *URA2* mRNA produced in the presence of 6-azauracil, which may reflect a particularly effective depletion in UTP, due to its inability to activate *URA1* and *URA3*.

### Organisation of the *URA2* promoter region

It was previously thought that *URA2* transcription starts some 70 nt ahead of the ATG start codon (Potier *et al*, 1990). This is clearly inconsistent with the recent report of an untranslated intron between positions -385 and -66 relatively to the ATG (Juneau *et al*, 2007). As shown in Figure 2, we confirmed the existence of this intron, although with different 3'-ends corresponding to positions -63, -54 or, in one case, +12. The corresponding DNA has a 5'-end consensus (-384GUAUGU-379), a canonical branch point box (-92UACUAC-84), and a -68UAG-65 at the 3'-splice site. An intron deletion did not detectably affect the expression of *pURA2::LacZ* (bottom line of Figure 2A).

Using RNA ligase mediated-RACE assays, we obtained and sequenced 13 cDNAs that, by construction, extended to position +188 of the *URA2* open reading frame and thus corresponded to the 5'-end domains of full-length mRNAs (Figure 2B). These 5'-ends defined six positions (-588A, -579T, -576A, -571A, -565A and -563A) clustered within an A-rich domain, matching to the initiator consensus ( $A_{rich}$ )<sub>5</sub>N(C/T)<sub>1</sub>A(A/T)NN( $A_{rich}$ )<sub>6</sub> deduced from a recent survey of yeast mRNAs (Zhang and Dietrich, 2005). A parallel study (Thiebaut *et al*, 2008) has identified a second initiator region (positions -686/-657, denoted by a grey arrow in Figure 2A) upstream of the one shown here to be used for full-length *URA2* mRNAs. This upstream initiator produces short unstable non-coding RNAs with the same transcriptional orientation as *URA2*. It is evidently not used to produce full-length *URA2* transcripts, as the predicted 5'-ends were not found in the corresponding cDNAs (Figure 2B). Five TCTT (UCUU) boxes, present immediately upstream of the



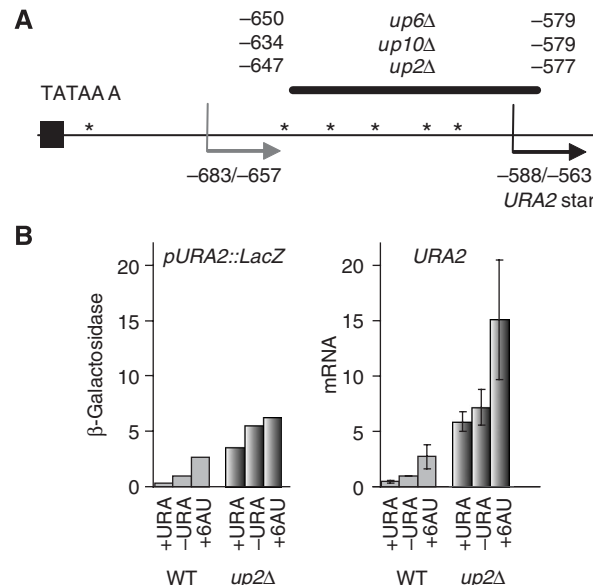
**Figure 2** Organisation of the *URA2* promoter region. **(A)** Deletion mutagenesis of promoter region. *LacZ* expression in BMA64-*URA3* (WT) transformed with pFL80 (*pURA2::LacZ*), pFL80-KBΔ (*KpnI*-*BsrGI* deletion), pFL80-BSΔ (*BsrGI*-*SalI* deletion) and pFL80-INTΔ (no intron). Strains were exponentially grown without (–URA) or with uracil (+URA), or with 200 mg/l of 6-azauracil (+6AU). β-galactosidase was expressed in arbitrary units (Miller, 1972), where 1.0 is the level measured in wild-type cells grown in CSM. Broken arrows correspond to the transcriptional initiator regions, shown in black for full-length *URA2* mRNA and in grey for short upstream untranslated RNAs (Thiebaut *et al*, 2008). A canonical TATAAA box and a cluster of TCTT (UCUU) motifs (stars) are also indicated (see Supplementary data). **(B)** Identification of the *URA2* 5'-end region and of an *URA2* intron. Thirteen cDNAs primed from an oligonucleotide corresponding to positions 167/188 of the *URA2* open reading frame were amplified by the RNA Ligase Mediated-RACE technique (Materials and methods). The six *URA2* start sites (underlined) fall between positions –589 and –563, containing the unique *SalI* site (*gtcgac*). Nine cDNAs were extracted and amplified from wild-type cells (BMA64-*URA3*) grown without uracil (–URA), or with 200 mg/l of 6-azauracil (+6AU), and four others were prepared from FL-M9 (*rpb1-L1397S*) grown without uracil (–URA). This also confirmed the existence of an intron between positions –385 and –66 (Juneau *et al*, 2007), but with 3'-borders corresponding to positions –63, –54 or, in one case, +12.

*URA2* initiator, form a cluster of binding sites recognised by Nab3, which belongs to the transcriptional terminator system operating on non-coding intergenic RNAs (Arigo *et al*, 2006; Thiebaut *et al*, 2006).

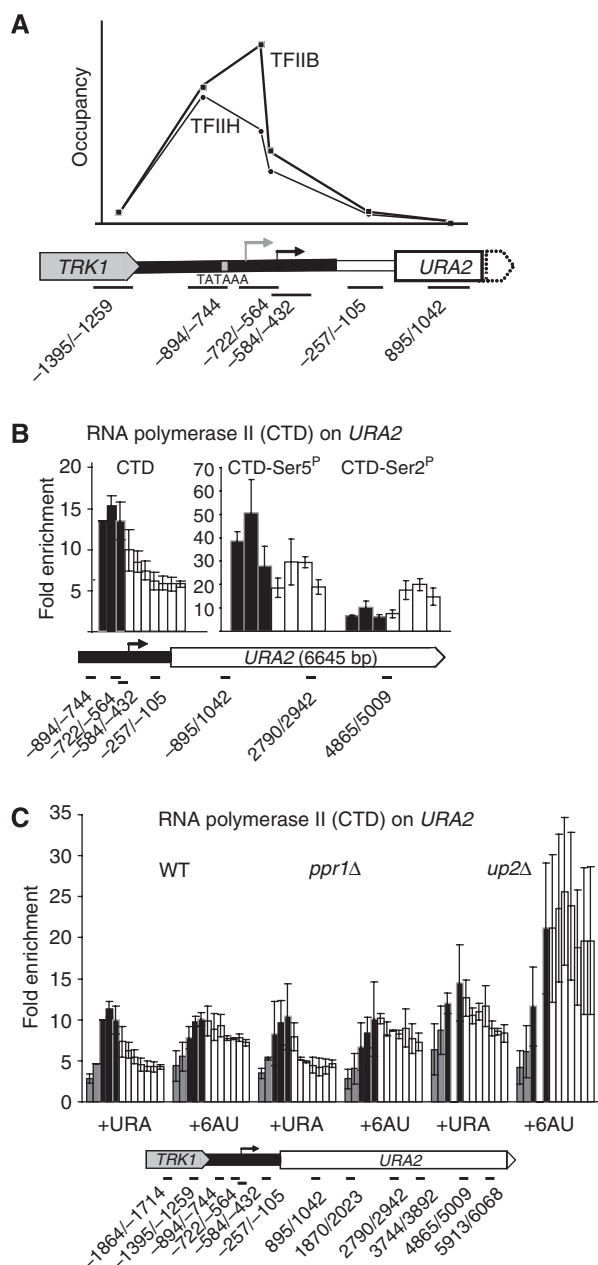
### The DNA upstream of the *URA2* initiator prevents the progression of RNA polymerase II

Partial deletions of the *TRK1-URA2* intergenic region were introduced in *pURA2::LacZ* reporters (pFL80 and pFL80-H2) and tested for *LacZ* activation under repressing (uracil) and derepressing (6-azauracil) conditions. Deleting the first 297 nt had no effect on *pURA2::LacZ*, but a *BsrGI/SalI* deletion lacking positions –934/–580 almost entirely blocked *pURA2::LacZ* expression (Figure 2A). The corresponding DNA is therefore critical for *URA2* transcription. It contains a TATAAA box (–776/–771) shown by Thiebaut *et al* (2008) to be needed for the synthesis of the *URA2* mRNA and of its short-lived upstream transcripts.

The *BsrGI/SalI* DNA was submitted to error-prone amplification and recombined into the *pURA2::HIS3* reporter (pFL83), producing plasmids that were selected for *HIS3* overexpression in the presence of uracil. No single-base mutation was isolated by this approach, but we obtained three very similar deletions *up2Δ* (–647/–577), *up6Δ* (–634/–579) or *up10Δ* (–650/–579) upstream of the *SalI* site (Figure 3A). These deletions overexpressed *LacZ* when subcloned into *pURA2::LacZ* reporters (pFL80-H2 and pFL80), and a chromosomal *up2Δ* mutant (FL51) constitutively expressed the *URA2* mRNA under-repressing conditions (Figure 3B). Moreover, *up2Δ* only acted in *cis*, as the



**Figure 3** The DNA upstream of the transcription start impairs *URA2* expression. **(A)** Schematic organisation of the *upΔ* mutations. Thick black line corresponds to the deleted DNA *up2Δ*, *up6Δ* and *up10Δ*. Stars indicate TCTT (UCUU) boxes. Broken arrows correspond to the transcriptional initiator regions as shown in Figure 2A. **(B)** Effect of *up2Δ* on *pURA2::LacZ* expression and *URA2* mRNA steady-state level. BMA64-*URA3* (WT) was transformed with pFL80 or pFL80-*up2Δ*. β-galactosidase was tested as shown in Figure 2A. *URA2* mRNA is expressed in arbitrary units (Miller, 1972), where 1.0 corresponds to wild-type cells grown in the absence of uracil (–URA). RT-PCR assays are based on three independent cultures of BMA64-*URA3* (WT) and FL51 (*up2Δ*), grown with (+URA) or without uracil (–URA), or containing 200 mg/l of 6-azauracil (+6AU).



**Figure 4** Distribution of TFIIB, TFIID and RNA polymerase II at the *URA2* locus. **(A)** TFIIB (Rad3-TAP) and TFIID (Sua7-TAP) were detected by ChIP assays as described in Materials and methods. Strains D712-10C (WT *RAD3::TAP*) and D714-5D (WT *SUA7::TAP*) were grown on SD + aa. The schematic organisation of the corresponding DNA is presented below. Broken arrows represent the *URA2* mRNA start sites and the upstream initiator, as in Figure 2A. **(B)** RNA polymerase II was immunoprecipitated with Dynabeads anti-mouse (Dyna Beads), using anti-CTD (8WG16), anti-Ser2<sup>P</sup> and anti-Ser5<sup>P</sup> antibodies (Covance). Strain GR44-11C (WT) was grown on SD + aa. A schematic map indicates the oligonucleotides used as primers. **(C)** RNA polymerase II occupancy in wild-type, *ppr1Δ* and *up2Δ*. Strains GR44-11C (WT), FL53 (*ppr1Δ*) and FL51 (*up2Δ*) were grown at 30 °C on SD + aa with 2 g/l of uracil (+URA) or exposed to 200 mg/l of 6-azauracil (+6AU). ChIP signals were detected with anti-CTD antibodies (8WG16). A schematic map represents the *URA2* gene, with the oligonucleotides used as primers.

promoter reporter (*pURA2::LacZ*) was not activated in the *up2Δ* host (data not shown).

The ChIP assays of Figure 4A show that the region upstream of the *URA2* start site is occupied by TFIIB and TFIID, two components of the RNA polymerase II pre-initiation complex. The corresponding RNA polymerase II signal was recognised by anti-CTD antibodies, and also by an anti-Rpb3::HA tag (see Figure 7D below), but not by antibodies raised against Ser2-phosphorylated CTD (Figure 4B). The CTD-Ser2<sup>P</sup> signal was only detected downstream of the *URA2* initiator region. As shown in Figure 4C, this also coincided with a drop in RNA polymerase II occupancy, observed in wild-type or *ppr1Δ* cells grown under repressing conditions.

*URA2* activation in wild-type or *ppr1Δ* cells exposed to 6-azauracil did not increase the RNA polymerase II signal detected upstream of the initiator region, but correlated with a full occupancy of the *URA2* open reading frame by RNA polymerase II (Figure 4C). Likewise, the *up2Δ* mutation produced a strong RNA polymerase II signal downstream of the initiator, even when grown under repressive conditions (Figure 4C). Thus, *URA2* activation is not due to an increased recruitment of RNA polymerase II (as in the case of *Ppr1*-dependent genes such as *URA1*; see Figure 1C) but results from an extended RNA polymerase II occupancy downstream of the *URA2* initiator region. Moreover, a relatively short DNA region, lost in *up2Δ* and located immediately upstream of the *URA2* initiator, impairs *URA2* transcription by preventing RNA polymerase II from progressing towards the *URA2* open reading frame.

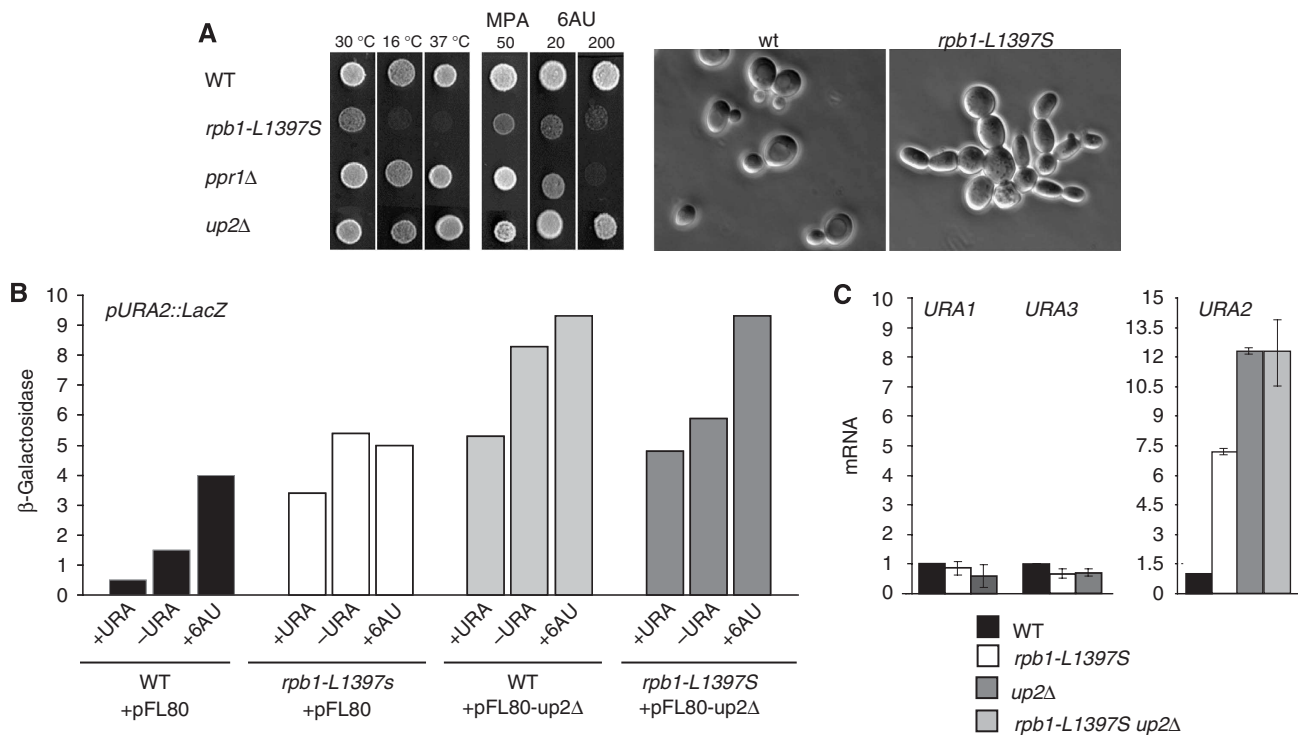
#### Mutants of the RNA polymerase II Switch 1 loop activate *URA2* expression

To search for *trans*-acting regulator(s) of *URA2*, mutations constitutively expressing the chromosomal *pURA2::HIS3* reporter of strain FL52 were selected after UV mutagenesis. We obtained four mutants (FL-M9, M10, M13 and M23) that were resistant to 3-aminotriazol on uracil-supplemented medium and also overexpressed the *pURA2::LacZ* reporter. They grew slowly at 30 °C, failed to grow at 16 °C and 37 °C and were somewhat sensitive to 6-azauracil (shown in Figure 5A for the FL-M9 strain). Meiotic tetrad analysis showed that these phenotypes co-segregated in a monogenic and recessive way, and complementation tests established that the corresponding mutations belonged to one and the same gene.

We then transformed FL-M9 with a yeast genomic library and obtained one plasmid restoring growth at 37 °C. The corresponding insert harboured *RPB1*, which encodes the largest subunit of RNA polymerase II, and three surrounding genes. Further subcloning showed that growth at 37 °C correlated with an intact *RPB1* gene, and complementation tests with the temperature-sensitive allele *rpb1-1* (Scafe *et al*, 1990) firmly established that all four mutants were due to *rpb1* mutations. *In vivo* gap repair (Rothstein, 1991) indicated that these mutations belonged to the *SuaI-PshAI* segment of *RPB1*. Finally, DNA sequencing revealed single amino-acid replacements corresponding to *rpb1-L1397S* (M9), *rpb1-S1401P* (M10) and *rpb1-F1402L* (M13 and M23), which were also present in the chromosomal DNA of the original mutant strains.

The above-mentioned data led to the rather unexpected conclusion that *URA2* activation occurs by *trans*-acting mutations altering RNA polymerase II itself, at the level of its Switch 1 loop fold. RT-PCR and *LacZ* assays confirmed that





**Figure 5** Properties of Switch 1 loop mutants. (A) GR44-11C (WT) and D711-13B (*rpb1-L1397S*) cells spotted on SD+aa or SD+aa supplemented with mycophenolate (50 mg/l) or 6-azauracil (20 and 200 mg/l). FL53 (*ppr1Δ*) and FL51 (*up2Δ*) were used as controls. Cells were microphotographed in exponential cultures grown in SD+aa at 30 °C. (B) Expression of *pURA2::LacZ*. β-galactosidase was assayed as shown in Figure 2A, in strains BMA64-URA3 (WT) and FL-M9-URA3 (*rpb1-L1397S*) transformed with the pFL80 or pFL80-*up2Δ* *pURA2::LacZ* reporter plasmids. (C) GR44-11C (WT) or YGH2 (WT), D711-13B (*rpb1-L1397S*), FL51 (*up2Δ*) and D876-10D (*rpb1-L1397S up2Δ*) were exponentially grown in CSM. Total RNA was extracted as described in Materials and methods. Steady-state levels of *URA1*, *URA2* and *URA3* mRNAs were quantified by RT-PCR and expressed in arbitrary units, where 1.0 corresponds to the wild-type level of *URA1* mRNA in CSM medium for *URA1* and *URA3*, and to the wild-type level of *URA2* mRNA in CSM.

*rpb1-L1397S* highly expressed *URA2* (Figure 5B and C). Moreover, an *rpb1-L1397S up2Δ* double-mutant had the same constitutive expression as *up2Δ* alone, and a *pURA2::LacZ* reporter bearing the *up2Δ* allele (plasmid pFL80-*up2Δ*) was expressed at the same level when tested in wild-type or in an *rpb1-L1397S* host strain (Figure 5B and C). In other words, *rpb1-L1397S* and *up2Δ* have epistatic effects on *URA2* expression, thus strongly suggesting that they are defective in the same mechanism downregulating *URA2* in response to uracil.

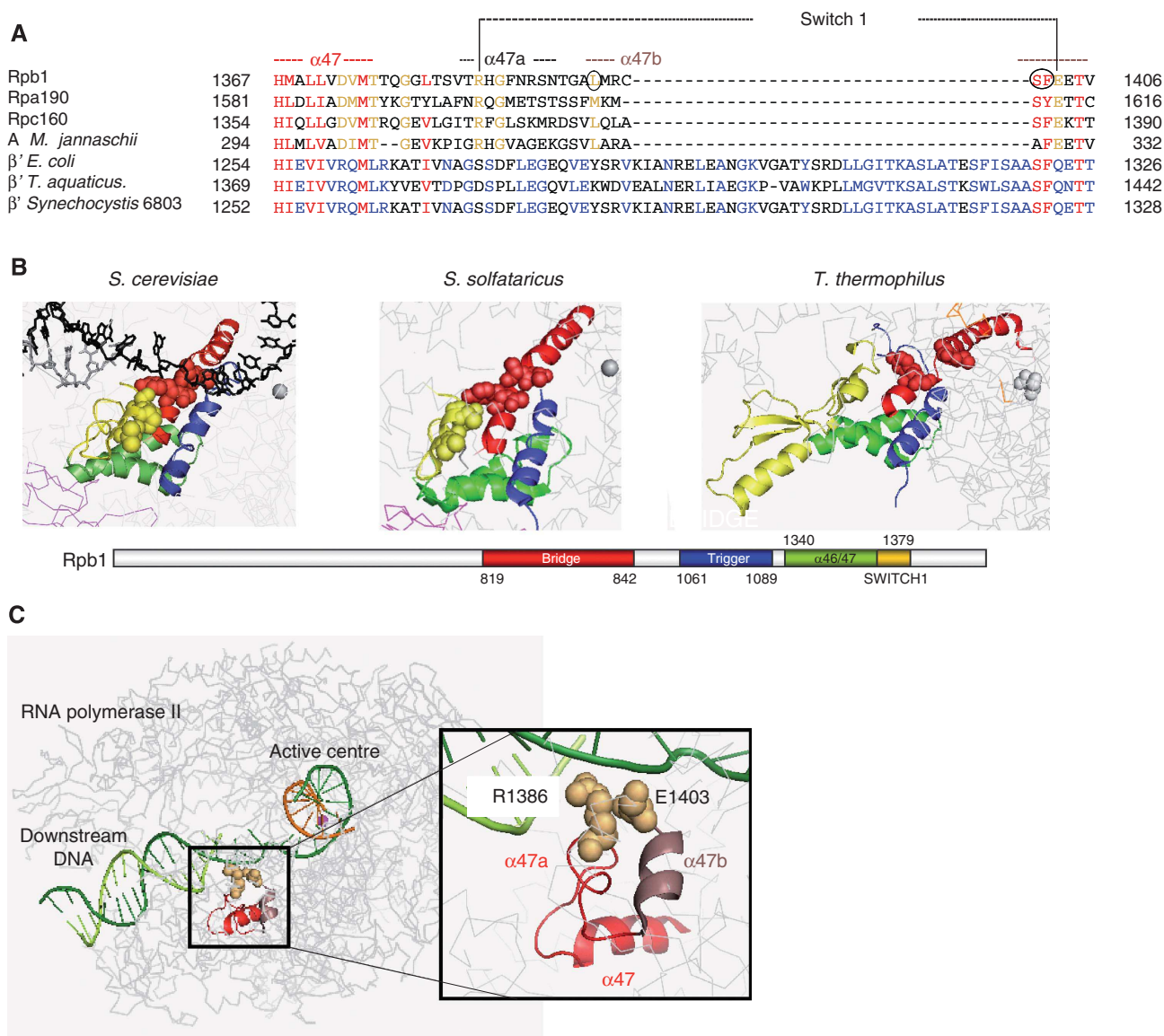
The six amino-acid segment occupied by L1397, S1401 and F1402 corresponds to Rpb1-α47b, one of the two α helices forming the Switch 1 loop domain of the RNA polymerase II active site (Gnatt *et al*, 2001). Moreover, the *rpb1-G1388V* allele, altering the Rpb1-α47a helix (Berroteran *et al*, 1994), also resulted in a high constitutive expression of *URA2* (data not shown). This high clustering was not anticipated in a UV mutagenesis, as the latter evidently affects the whole yeast genome. Our genetic screen was presumably not saturating, and we cannot exclude that *URA2* activation may, for some reason, be a general property of partly defective RNA polymerase II mutants. We thus tested the *pURA2::LacZ* reporter (pFL80) in 10 other slow-growing mutations partly impairing Rpb1 (*rpb1-G1437D*, *rpb1-E1351K* and *rpb1-H1367D*) or Rpb2 (*rpb2-R857K*, *rpb2-E836A*, *rpb2-D978A*, *rpb2-P1018S* and *rpb2-G1142D*) or lacking the non-essential subunits Rpb4 (*rpb4Δ*) or Rpb9 (*rpb9Δ*) (Woychik and Young, 1989; Scafe *et al*, 1990; Woychik *et al*, 1991). Except for a modest

derepressing effect of *rpb2-E836A* and *rpb2-R857K*, their β-galactosidase activity was equal to or lower than the wild-type control (data not shown), thus strongly suggesting that *URA2* activation, at least to a large extent, is primarily due to changes in the Switch 1 loop.

The Switch1 loop is highly conserved in all eukaryotic and archaeal RNA polymerases (Figure 6A). Together with the α25 Bridge helix, α36 Trigger helix and the α46/47 Loop of Rpb1, it forms an identical fold in the RNA polymerase of *Sulfolobus solfataricus* (Hirata *et al*, 2008) and in yeast RNA polymerase II (Gnatt *et al*, 2001). This fold wraps the downstream end of the transcription bubble and holds the DNA template strand by the invariant R1386-E1403 Switch 1 dipole (Figure 6B and C). As discussed elsewhere (Zaros *et al*, 2007), this fold is stabilised by the C-end of the Rpb5 subunit, itself strongly conserved from archaea to eukaryotes. Bacterial RNA polymerases, on the other hand, have no Rpb5 and their Switch 1 loop is considerably extended but nevertheless adopts a very similar spatial orientation (Vassilyev *et al*, 2007).

#### ***rpb1-L1397S* enhances RNA polymerase II occupancy downstream of the *URA2* initiator**

Consistent with the partial growth defects of *rpb1-L1397S*, a twofold reduction in RNA polymerase II occupancy was seen for *rpb1-L1397S* in single-gene ChIP assays at *ADH1* or *PYK1* (Figure 7A). This extended to the whole genome, with an average occupancy level of 62 % ( $R^2 = 0.80$ ) relatively to wild



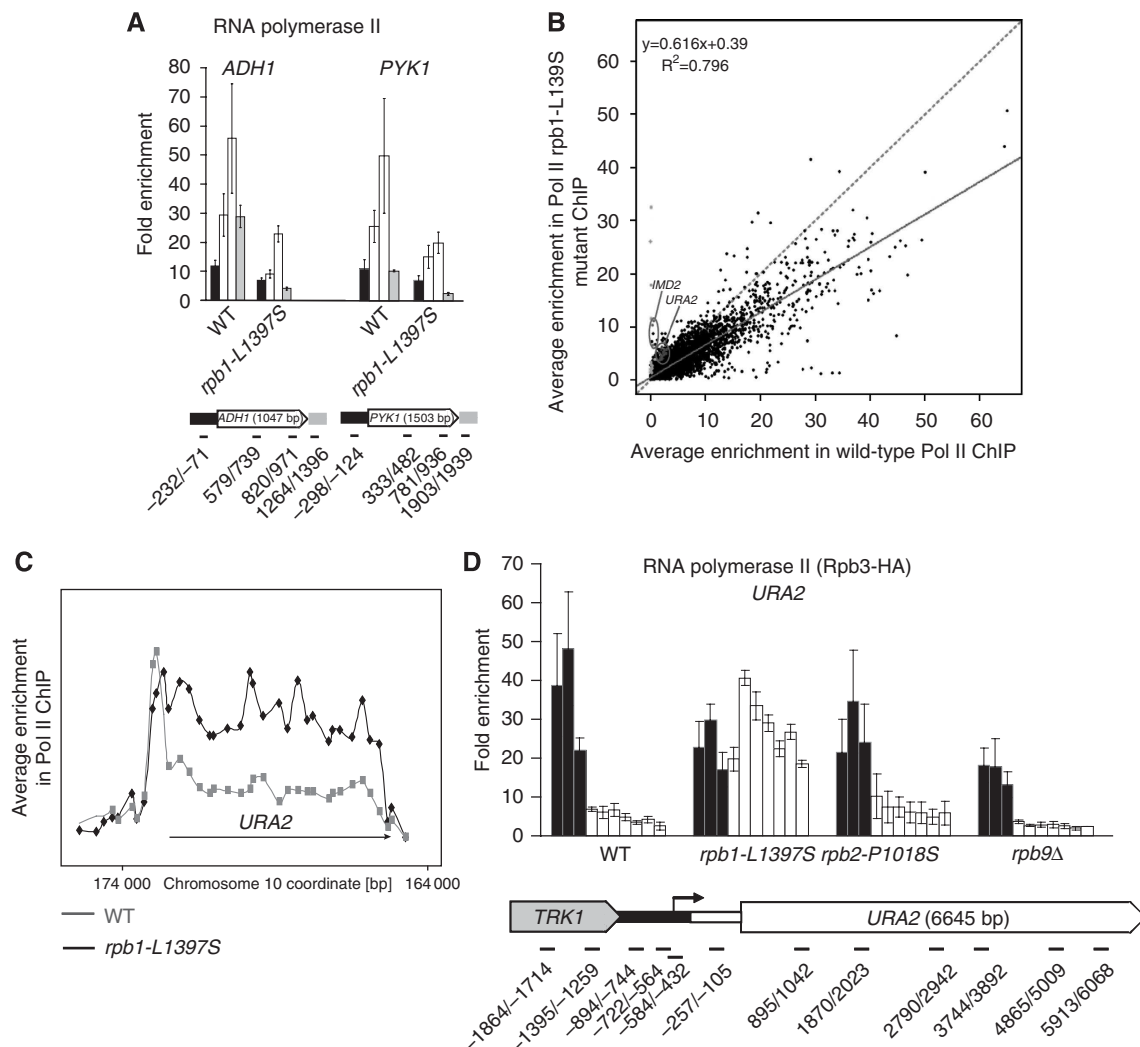
**Figure 6** Organisation of the Switch 1 loop in RNA polymerase II. (A) Sequence alignments with yeast, archaeal and bacterial subunits. The Switch 1 loop domains (Gnatt *et al*, 2001) of Rpb1, Rpa190 and Rpc160 (*S. cerevisiae*) were aligned with the corresponding region of *M. jannaschii* (subunit A) and of the β' subunits of *E. coli*, *Thermus aquaticus* and *Synechocystis* 6803. Amino-acid conservation is shown in red (all species), gold (Eukaryotes and Archaea) and blue (Prokaryotes only). Circles denote the mutated amino acids of *rpb1-G1388V*, *rpb1-L1397S*, *rpb1-S1401P* and *rpb1-F1402L*. (B) Spatial organisation of the Bridge, Trigger and Switch 1 fold in *S. cerevisiae*, *S. solfataricus* and *T. thermophilus* on the basis of PDB files 2NVZ, 2PMZ and 205I, respectively. These domains are shown on the Rpb1 sequence, with the same colour code and with numbers indicating the corresponding amino-acid positions. (C) Organisation of the Switch 1 loop in RNA polymerase II, using a different orientation, and showing the whole 10-subunit structure of RNA polymerase II (without Rpb4 and Rpb7). The template and non-template DNA strands and the nascent RNA are in dark green, light green and orange, respectively. A box shows the details of Switch 1 loop, underscoring the DNA-binding positions R1386 and E1403 by space-filling.

type (Figure 7B). As expected, RNA polymerase II occupancy was increased on the *URA2* open reading frame. Moreover, this increased occupancy was only seen downstream of the *URA2* initiator region (Figure 7C). Similar results were obtained in ChIP assays done in *RPB3::3HA rpb1-L1397S* cells, showing that this effect is independent of the precipitating antigen (Figure 7D). In contrast, two RNA polymerase II alleles (*rpb2-P1018S* and *rpb9Δ*) with partial growth defects similar to those of *rpb1-L1397S* significantly reduced RNA polymerase II signal, upstream and downstream of the initiator region. Taken together, these data clearly show that *rpb1-L1397S* activates *URA2* by extending RNA polymerase II

occupancy downstream of the initiator region, thus recapitulating the effect seen above for *up2Δ* or for wild-type and *prr1Δ* cells exposed to 6-azauracil.

#### ***rpb1-L1397S* extends RNA polymerase II downstream of *IMD2*, *IMD3* and *URA8***

We have seen above that *rpb1-L1397S* generally reduced RNA polymerase II occupancy, except for *URA2*. This was also consistent with genome-wide transcriptome assays, showing a moderate but significant overexpression of *URA2* in *rpb1-L1397S*, compared with an isogenic wild-type control (Figure 8A). Furthermore, these transcriptome and genome-



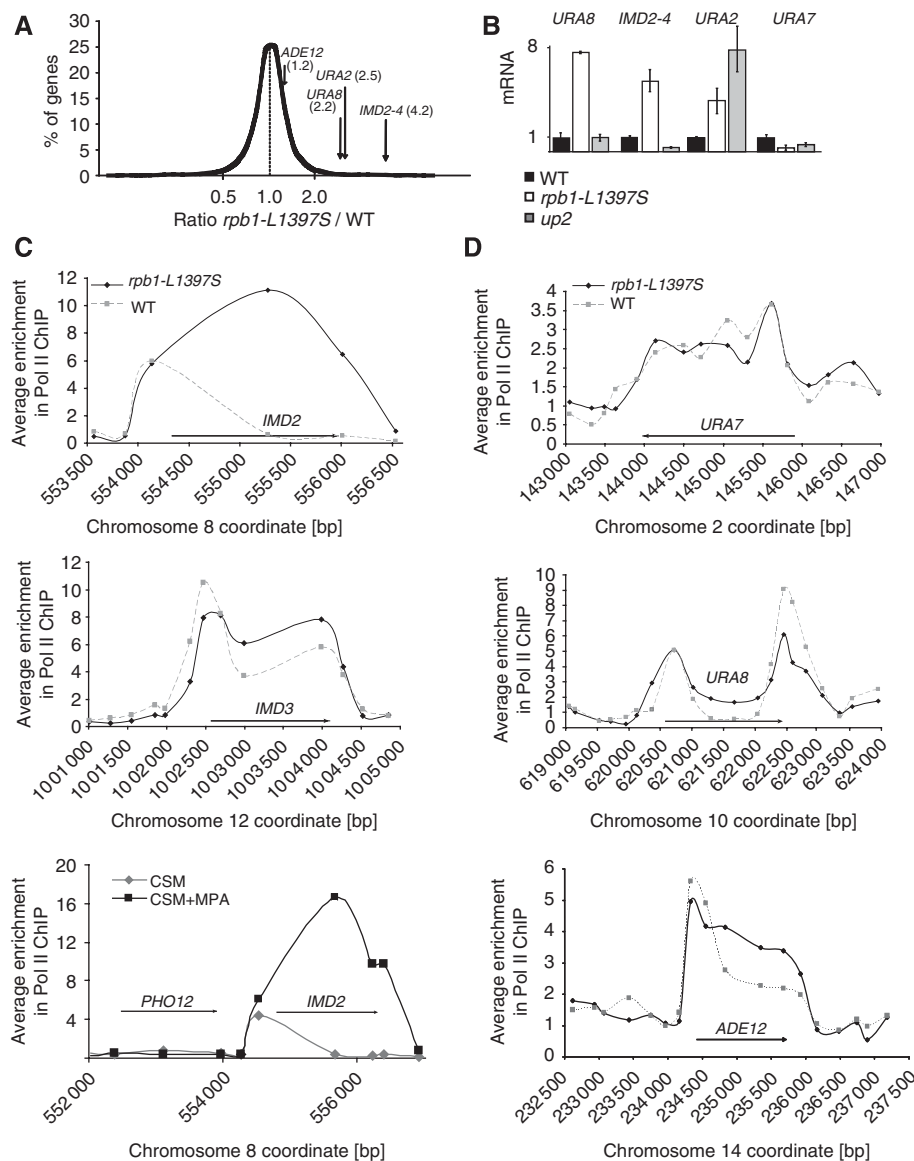
**Figure 7** RNA polymerase II occupancy in *rpb1-L1397S*. **(A)** RNA polymerase II occupancy at *ADH1* and *PYK1* in D535-4D (*RPB3::HA*) and D535-9D (*rpb1-L1397S RPB3::HA*). Cells were grown in SD + aa at 30 °C. RNA polymerase II was immunoprecipitated with anti-HA antibodies. A schematic map represents the oligonucleotide pairs used as primers. Position +1 corresponds to the initiator ATG. **(B)** Genome-wide ChIP assays. Three independent cultures of YGH2 (wild-type) and D711-13B (*rpb1-L1397S*) were grown in YPD at 30 °C. Chromatin was extracted and analysed as described previously (Ghavi-Helm *et al*, 2008). **(C)** Distribution of RNA polymerase II at the *URA2* locus. Data were extracted from genome-wide analysis shown in Figure 7B above. The wild-type and *rpb1-L1397S* profiles are shown in grey and black, respectively. **(D)** Effect of *rpb1-L1397S*, *rpb2-P1018S* and *rpb9Δ* on RNA polymerase II occupancy at *URA2*. Experimental conditions were as shown in Figure 7A. The strains used were D535-4D (*RPB3::HA*), D535-9D (*rpb1-L1397S RPB3::HA*), YMW304-8C (*rpb2-P1018S RPB3::HA*) and YWM305-2D (*rpb9Δ RPB3::HA*).

wide RNA polymerase II occupancy assays identified two other genes, *IMD2* (IMP dehydrogenase) and *URA8* (CTP synthase), which were both overexpressed and over-occupied by RNA polymerase II in *rpb1-L1397S*. A few additional genes were also overexpressed in the *rpb1-L1397S* transcriptome, but with no higher RNA polymerase II occupancy of their open reading frames (data not shown). Conversely, *rpb1-L1397S* enhanced RNA polymerase II at *ADE12* (adenylosuccinate synthase), but had no detectable effect on its mRNA level.

*Saccharomyces cerevisiae* has two CTP synthases (*URA7* and *URA8*) and three active IMP dehydrogenase genes (*IMD2*, *IMD3* and *IMD4*), with *IMD1* as a pseudo-gene inactivated by a frame-shift mutation in most laboratory strains (Ozier-Kalogeropoulos *et al*, 1994; Nadkarni *et al*, 1995; Hyle *et al*, 2003). *URA7* and *URA8* are sufficiently different to be distinguished by individual RT-PCR

assays, which clearly showed that *rpb1-L1397S* activated only *URA8* (Figure 8B). RT-PCR assays also confirmed the overexpression of the IMP dehydrogenase mRNA in *rpb1-L1397S* but did not clearly distinguish between *IMD2*, *IMD3* and *IMD4*, due to their very similar nucleotide sequences. However, our genome-wide ChIP assays readily discriminated between these three genes. This revealed that *rpb1-L1397S* strongly activated *IMD2*, with some effect on *IMD3* (Figure 8C) and no effect at all on *IMD4* (not shown). We considered the possibility that the activation of CTP synthase and IMP dehydrogenase genes may be an indirect effect of *URA2* overexpression, reflecting an increased cellular pool of UTP. However, the activation of *URA2* in *up2Δ* was not accompanied by a parallel increase in *URA8* or IMP dehydrogenase mRNA, thus strongly arguing for a direct activating effect of *rpb1-L1397S* (Figure 8B).





**Figure 8** Effect of *rpb1-1397S* on *IMD2*, *IMD3*, *ADE12*, *URA7* and *URA8*. (A) Enrichment of RNA polymerase II transcripts in D535-9D (*rpb1-L1397S*) compared with an isogenic wild-type obtained by complementation with pFL36-RPB1. The histogram represents median percentile ranks of Cy3/Cy5 fluorescence ratios. Data are based on three independent cultures of each strain. The *URA2*, *URA8*, *ADE12* and *IMD2-4* transcripts are individualised by arrows. (B) RT-PCR of individual mRNAs. GR44-11C (WT), D711-13B (*rpb1-L1397S*) and FL51 (*up2Δ*) were exponentially grown in SD + aa. Total RNA was extracted as described in Materials and Methods. Individual mRNAs were quantified by RT-PCR and expressed in arbitrary units, where 1.0 corresponds to the wild-type level. (C) RNA polymerase II occupancy at *IMD2* and *IMD3*. Data were extracted from the genome-wide analysis shown above (Figure 7B) except for *IMD2* in wild-type cells grown in CSM + MPA, taken from Ghavi-Helm *et al* (2008). (D) RNA polymerase II occupancy at *URA7*, *URA8* and *ADE12*. Data were extracted from the genome-wide analysis shown above (Figure 7B).

Previous studies have shown that *IMD2* is controlled by an initiation switch between short transcripts, synthesised from an upstream initiator, and full-length mRNAs starting from a downstream initiator (Escobar-Henriques *et al*, 2003; Steinmetz *et al*, 2006; Kopcewicz *et al*, 2007; Jenks *et al*, 2008). RNA polymerase II is confined to the upstream initiator when wild-type cells are grown under repressing conditions (Steinmetz *et al*, 2006), but extended to the entire *IMD2* gene in *rpb1-L1397S* or in derepressed wild-type cells that have been exposed to mycophenolate (Figure 8C). In *URA8* and perhaps in *ADE12*, Thiebaut *et al* (2008) have provided evidence for a transcriptional switch reminiscent of the *IMD2* case. Again, RNA polymerase II was mostly restricted to the upstream part of the gene in wild-type cells

grown under repressing conditions, but extended to the entire *URA8* open reading frame in *rpb1-L1397S* (Figure 8C). A similar effect may apply to *ADE12*, encoding the first step of the ATP biosynthetic pathway, where *rpb1-L1397S* moderately increased RNA polymerase II occupancy downstream of the promoter region (Figure 8D).

## Discussion

Previous studies have shown that *URA1* or *URA3* are specifically activated by Ppr1, a Zn activator of the Gal4-type which responds to increased concentration of the orotate and/or dihydro-orotate precursors of UTP (Loison *et al*, 1980; Losson and Lacroute, 1981; Flynn and Reece, 1999).

This leads to an increased RNA polymerase II occupancy at the *URA1* and *URA3* promoters, lost in the *ppr1Δ*-null allele, indicating a 'classical' mode of gene-specific activation on the basis of a more effective recruitment of RNA polymerase II by its pre-initiation complex. *URA2*, which encodes the main rate-limiting enzyme of UTP biosynthesis, is activated when UTP is depleted by 6-azauracil addition. However, the *ppr1Δ*-null allele is fully competent for *URA2* transcription. Moreover, and in contrast to *URA1* or *URA3*, *URA2* activation does not change the amount of RNA polymerase II residing at or upstream of the *URA2* initiator region, but operates instead by extending RNA polymerase II occupancy to the *URA2* open reading frame.

Three *cis*-acting mutations constitutively expressing *URA2* were selected in this study. They corresponded to very similar deletions (*up2Δ*, *up6Δ* and *up10Δ*) removing a 55–70 nt DNA region upstream of the *URA2* initiator. The *up2Δ* mutation was studied in more detail. It extended RNA polymerase II to the entire *URA2* open reading frame, with no change in the RNA polymerase II signal upstream of the *URA2* initiator, thus recapitulating the pattern seen in UTP-depleted wild-type or *ppr1Δ* cells. Further studies (Thiebaut *et al*, 2008) have shown that the corresponding DNA is transcribed from an initiator region located some 100 nt upstream of the main *URA2* mRNA 5'-ends identified here, producing unstable RNAs with the same transcriptional orientation as *URA2*, which were no longer detected in *up2Δ*. Taken together, these data suggest that *URA2* may be regulated by an attenuation mechanism related to the one recently described for *IMD2* (Escobar-Henriques *et al*, 2003; Davis and Ares, 2006; Steinmetz *et al*, 2006; Kopcewicz *et al*, 2007; Jenks *et al*, 2008).

A genome-wide selection for *trans*-acting mutations constitutively expressing *URA2* yielded three tightly clustered *rpb1* mutations (*rpb1-L1397S*, *rpb1-S1401P* and *rpb1-F1402L*), corresponding to the Rpb1 $\alpha$ -47b helix of the Switch 1 loop. *rpb1-G1388V*, which was initially selected for its ability to alter transcription initiation in favour of downstream sites (Berroteran *et al*, 1994) and belongs to the Rpb1 $\alpha$ -47a helix, also activates *URA2* (see also Thiebaut *et al*, 2008). In contrast, other partly defective RNA polymerase II mutations failed to derepress *URA2*, indicating that *URA2* activation is by no means a general property of RNA polymerase II mutants. Similar to *up2Δ*, *rpb1-L1397S* activates *URA2* by extending RNA polymerase II occupancy to the entire open reading frame, with no change in the signal upstream of the *URA2* initiator, thereby reproducing the effect seen in *up2Δ* and in UTP-depleted wild-type cells. Moreover, *up2Δ* and *rpb1-L1397S* had epistatic effects on *URA2* expression, indicating that *rpb1-L1397S* is defective in the attenuation mechanism lost in *up2Δ* itself. Finally, *rpb1-L1397S* and wild-type *URA2* mRNAs had similar 5'-ends, as determined by 5'-RACE assays, and therefore use the same transcription initiator.

The above-mentioned data indicate that changes in the Switch 1 loop specifically activate *URA2*, despite their adverse effects on growth and their reduced genome-wide RNA polymerase II occupancy. Moreover, we were unable to identify mutants encoding a specific regulator of *URA2*, although it remained possible that such mutants are lethal because they perturb some essential cellular process. The Switch 1 loop forms an invariant R1386( $\alpha$ 47a)-E1403( $\alpha$ 47b)

dipole, holding the DNA template strand at positions +2/+3 downstream of the catalytic Mg<sup>2+</sup> (Gnatt *et al*, 2001). This domain is highly conserved among eukaryotic and archaeal RNA polymerases (Cramer *et al*, 2001; Hirata *et al*, 2008) and a related structure exists in the bacterial enzyme, where it also binds DNA at positions +2/+3 downstream of the catalytic Mg<sup>2+</sup> (Vassilyev *et al*, 2007; Zaros *et al*, 2007).

The relation of this domain to the entry route of NTPs in RNA polymerase II is unclear (Landick, 2005). One possibility is that NTPs reach the catalytic site of RNA polymerase II by its funnel-shaped pore (Cramer *et al*, 2001; Kettenberger *et al*, 2003; Westover *et al*, 2004; Landick, 2005), but it has also been suggested that NTPs enter through the large DNA channel, transiently binding the DNA template at nucleotides +2, +3 and possibly +4 (Gong *et al*, 2005), which implies that NTPs pass by the Switch 1 loop before reaching the catalytic Mg<sup>2+</sup>. Hence, the Switch 1 loop may act as an NTP-sensing module, enabling RNA polymerase II to adopt a processive mode of elongation when enough NTP is available. This would optimise the efficiency of transcription, which may account for the cold- and heat-sensitive defects of *rpb1-L1397S*, *rpb1-S1401P* and *rpb1-F1402L* and their increased sensitivity to 6-azauracil and mycophenolate.

Strikingly, *rpb1-L1397S* also extended RNA polymerase II occupancy downstream of the *IMD2/IMD3* (IMP dehydrogenase) and *URA8* (CTP synthase) promoters, with some evidence for a similar effect on *ADE12* (adenylosuccinate synthase). This makes biological sense, as the corresponding genes encode rate-limiting enzymes in the *de novo* synthesis of GTP, CTP and ATP. This suggests the rather non-canonical view that nucleoside triphosphate shortage or RNA polymerase II Switch 1 loop mutations (assumed here to mimic that shortage) specifically activate the expression of genes that are themselves critical for the *de novo* synthesis of NTPs. In other words, the transcriptional response to NTP shortage would not rely on dedicated activators or repressors (which have been vainly searched for in the case of *URA2* and *IMD2*) but would be mediated by an NTP-sensing mechanism built in the structure of RNA polymerase II itself.

In *IMD2*, there is good evidence that RNA polymerase II molecules are recruited on a common pre-initiation complex, followed by a start site switch leading to the alternative production of upstream transcripts ended by Nab3/Nrd1-dependent termination or full-length mRNAs transcribed from a downstream initiator region. Start site selection is presumably dictated by the nucleotide composition of the surrounding DNA, with upstream transcription starting at Gs, whereas *IMD2* transcription starts in a G-poor DNA region and is therefore favoured by low GTP pools (Escobar-Henriques *et al*, 2003; Davis and Ares, 2006; Steinmetz *et al*, 2006; Kopcewicz *et al*, 2007; Jenks *et al*, 2008). The *URA2* start choice, however, depends on a different mechanism, as upstream initiation does not occur preferentially at T's and is not sensitive to the abundance of uracil (Thiebaut *et al*, 2008). Conversely, we found that 3 out of 10 *URA2* cDNAs produced under activating conditions started with a T, which argues against the simple idea of a selection driven by the nature of the starting nucleotide. Nevertheless, the upstream initiator is followed by an ~60 nt T-rich domain, continued by 40 nt that are conspicuously poor in T's and harbour the *URA2* mRNA 5'-ends. This pattern is conserved among all *Saccharomyces* species sequenced to data (see

Supplementary data), and we would be surprised if it did not provide some form of UTP-sensing. Further studies are clearly needed to better understand how RNA polymerase II responds to NTP shortage, and how this response is ultimately converted into a specific control of the *IMD2*, *URA2* and *URA8* genes.

## Materials and methods

### Plasmids

Newly constructed plasmids are listed in Supplementary data. Plasmids pFL35-II, pFL36-CII (*CEN6 LEU2*) and pFL39-CII (*CEN6 TRP1*), corresponding to pFL35, pFL36 and pFL39 (Bonneaud *et al*, 1991), with an extended polylinker of 90 nucleotides (Supplementary data). pFL80 and pFL82 are *pURA2::LacZ* *pURA2* reporters derived from pFL39-CII and pFL36-CII, respectively. They harbour a *KpnI*–*Bam*HI–*NotI* *pURA2::LacZ* cassette of 4392 nt. The *KpnI*–*Bam*HI segment (1254 nt) is formed by the *TRK1*–*URA2* intergenic region, preceded by *GGTACCAGCACACGCTCTAA* (the *KpnI* site is in italics) and followed by *ATGGATCC* (*Bam*HI). It is cloned in-frame to the initiator ATG of a *Bam*HI–*NotI* *LacZ* cassette. pFL83 is pFL80 where the *LacZ* open reading frame has been replaced by a *Bam*HI–*NotI* *HIS3* (*S. cerevisiae*) cassette. As pFL80 poorly complemented *trp1* mutant strains, which overexpressed the *pURA2* promoter (due to some transcriptional interference between *LacZ* and *TRP1*), we also used pFL80-H2, a *pURA2::LacZ* reporter with a lower *LacZ* expression. This plasmid was selected after random insertion of genomic *Hind*III fragments in the unique *Hind*III site downstream of *LacZ* (Supplementary data). The *Hind*III fragment of pFL80-H2 corresponded to positions 7679–8251 of the *TEL1* gene and was inserted in a transcriptional orientation opposite to the one of *TRP1*.

Plasmids pFL80-KB, pFL80-BS and pFL80-CS correspond to *KpnI*/BsrGI, *BsrGI*/SalI and *Clal*/SalI deletions of pFL80, respectively. pFL80-INTA was precisely deleted of the DNA comprised between nucleotides –384 to –63, where 1 corresponds to the *URA2* start codon. The pFL83-upΔ2, -up6Δ and -up10Δ plasmids were obtained by error-prone amplification of the *KpnI*–*Bam*HI cassette of pFL83, which was recombined into the original vector by co-transformation with linearised pFL83 DNA deleted of its *BsrGI*–*SalI* fragment. Mutant plasmids were selected on 3-aminotriazol in the presence of uracil, amplified in *Escherichia coli*, retransformed and verified for their overexpression. The corresponding upΔ alleles were transferred to pFL80 by *KpnI*–*Bam*HI subcloning.

### Yeast strains

Yeast mutants are listed as Supplementary data. Plasmid shuffle assays were done in the presence of 5-fluoro-orotic acid to counter-select *URA3* plasmids (Boeke *et al*, 1984). Except when stated otherwise, they were constructed in BMA64-1A and BMA64-URA3, a *URA3*<sup>+</sup> revertant of BMA64-1A (Baudin-Baillieu *et al*, 1997). Yeasts were grown on SD + aa, corresponding to the standard synthetic dextrose medium supplemented with histidine, tryptophan and adenine sulphate (20 mg/l), leucine and lysine (30 mg/l). In β-galactosidase assays, SD + aa was replaced by the complete synthetic medium CSM (manufactured by BIO101) supplemented with a complete set of amino acids but lacking uracil. CSM supplemented with 2 g/l of uracil (CSM + U) fully represses *URA2*, whereas adding 6-azauracil at 20 mg/l (CSM + 6AU) fully activates *URA2*.

FL50 was obtained by inserting the integrative plasmid pFL81 (*TRP1 pURA2::HIS3*) at the *EcoRV* site of the *trp1-1* allele in W303-1B. When grown in the presence of uracil, this strain has an increased sensitivity to 3-aminotriazol, due to the repression of its *pURA2::HIS3* reporter. FL52 is FL50 transformed with pFL82 (*LEU2 pURA2::LacZ*). FL-M9, M10, M13 and M23 are FL52 mutants selected from UV-irradiated cells (about 1% survival) plated for 3 days at about  $4 \times 10^6$  cells on leucine omission plates with uracil (2 g/l) and 50 or 100 mM of 3-aminotriazol. β-galactosidase was assayed in a loop of cells put on Whatman paper soaked in 2 ml of Z buffer (Miller, 1972) containing X-Gal (1 g/l) and a drop of zymolase. Dark blue clones were tested again, using freshly grown cells suspended in 0.6 ml of Z buffer and vortexed for 1 min with an equal volume of glass beads. Cells were refrigerated in an ice bath, treated with a second round of glass bead extraction, pelleted by centrifugation at 14 000 r.p.m. and tested for β-galactosidase after

10 min at 37 °C. Mutant strains with an at least fivefold increase in β-galactosidase were retained for further analysis. A fifth mutant was a *fur1* allele partly defective for UMP pyrophosphorylase, impairing the metabolic conversion of uracil into UMP.

FL51, with the *up2Δ* allele integrated on the chromosome by homologous recombination, was constructed by subcloning the *KpnI*–*Bam*HI fragment of pFL81-up2Δ into pFL35-II. The resulting plasmid was linearised at the *BsmI* site of the *pURA2* promoter region, to direct its integration by homologous recombination with the *pURA2* region of BMA64-URA3. The corresponding transformants harboured the *TRP1* cassette flanked by wild-type and *up2Δ* tandem copies of *pURA2*. Tryptophan auxotrophic clones spontaneously occurred by recombination between these *pURA2* copies and were selected by nystatin enrichment (Snow, 1966), yielding single copies of *pURA2*. Among them, *up2Δ* mutants were identified by their weak resistance to 5-fluorouracil ( $10^{-5}$  M), and were further checked by PCR amplification.

### ChIP and genome-wide ChIP-chip assays

Chromatin immunoprecipitation assays were as described elsewhere (Ghavi-Helm *et al*, 2008). All experiments were performed on three independent cultures of 100 ml, harvested at an OD<sub>600</sub> of 0.3–0.5. Cells were grown in SD + aa with or without uracil, or UTP-depleted by adding 200 mg/l of 6-azauracil to log-phase cells grown in SD + aa and grown for three additional doubling times. RNA polymerase II was immunoprecipitated with Dynabeads anti-mouse IgG (DynaL Biotech), using anti-CTD antibodies (8WG16), or anti-hemagglutinin A antibodies (12CA5) in the case of Rpb3-HA tagged strains. Sua7-TAP and Rad3-TAP proteins were directly immunoprecipitated on Dynabeads anti-mouse IgG. Phosphorylated variants of the RNA polymerase CTD domain were immunoprecipitated with Dynabeads anti-mouse IgM, using anti-Ser2<sup>P</sup> (H5, Covance) and anti-Ser5<sup>P</sup> (H14, Covance) antibodies. ChIP signals were calculated by the immunoprecipitation/input signal. The value 1.0 was arbitrarily given to the reference signal provided by amplifying the *GAL1* gene. Genome-wide ChIP assays, on the basis of three independent cultures in YPD or CSM medium for mycophenolate experiments (10 mg/l) medium, were described elsewhere (Harismendy *et al*, 2003) and were based on DNA arrays with over 40 000 oligonucleotide probes covering 12 Mb of the yeast genome (Ghavi-Helm *et al*, 2008).

### RNA and transcriptome assays

Total RNA was extracted from three independent, exponential cultures with hot phenol and reverse-transcribed using 1 μg of total RNA, Super-Script II reverse transcriptase (Invitrogen) and random hexamers as primers. Controls without reverse transcriptase showed negligible levels of DNA contamination. DNA was quantified by real-time PCR amplification (Applied Biosystems, System SDS Software) using primers listed in Supplementary data. mRNA levels are calculated as a ratio of measured mRNA and *ACT1* mRNA. The 5'-ends of *URA2* mRNAs were mapped by sequencing cDNAs obtained by the RNA ligase mediated-RACE technique (using an RLM-RACE kit from Ambion), according to the protocol provided by the manufacturer. Yeast micro-arrays were probed against total RNA extracted from three independent cultures of strains BMA64-URA3 (wild type) and FL-M9 (*rpb1-L1397S*) grown on SD + aa at 30 °C, to an OD<sub>600</sub> of 0.55. cDNA synthesis, purification, indirect labelling, microarray hybridisation, scanning and analysis were done according to standard protocols (Soutourina *et al*, 2006), with two independent hybridisations for each RNA batch.

### Supplementary data

Supplementary data are available at *The EMBO Journal* Online (<http://www.embojournal.org>).

## Acknowledgements

We thank Fred Beckouët, Antonin Morillon, Bertrand Séraphin and Michel Werner for strains, plasmids and helpful discussions. We are especially thankful to Alain Jacquier and Domenico Libri for communicating unpublished data before publication. This work was partly funded by the Ligue pour la Recherche sur le Cancer (FL). DD and MK benefited from fellowships of the French Association pour la Recherche sur le Cancer and Fondation pour la Recherche Médicale.

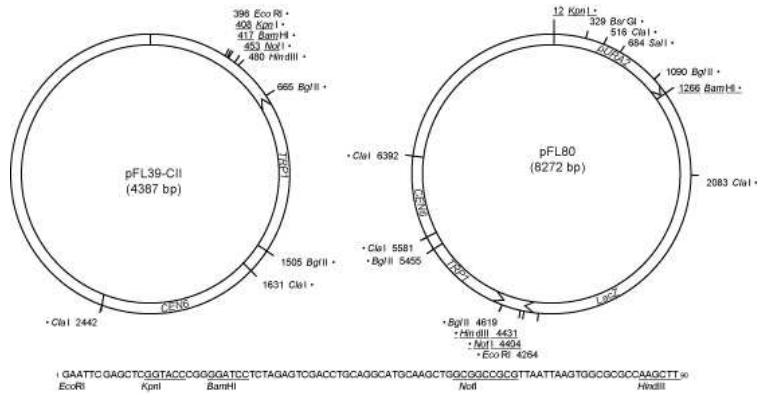
## References

- Arigo JT, Carroll KL, Ames JM, Corden JL (2006) Regulation of yeast NR1 expression by premature transcription termination. *Mol Cell* **21**: 641–651
- Baudin-Baillieu A, Guillemet E, Cullin C, Lacroute F (1997) Construction of a yeast strain deleted for the TRP1 promoter and coding region that enhances the efficiency of the polymerase chain reaction-disruption method. *Yeast* **13**: 353–356
- Berteroan RW, Ware DE, Hampsey M (1994) The *sua8* suppressors of *Saccharomyces cerevisiae* encode replacements of conserved residues within the largest subunit of RNA polymerase II and affect transcription start site selection similarly to *sua7* (TFIIB) suppressors. *Mol Cell Biol* **14**: 226–237
- Boeke JD, Lacroute F, Fink GR (1984) A positive selection for mutants lacking orotidine-5'-phosphate decarboxylase activity in yeast: 5-fluoro-orotic acid resistance. *Mol Gen Genet* **197**: 345–346
- Bonneaud N, Ozier-Kalogeropoulos O, Li G, Labouesse M, Minvielle-Sebastia L, Lacroute F (1991) A family of low and high copy replicative, integrative and single-stranded *S. cerevisiae*/E. coli shuttle vectors. *Yeast* **7**: 609–615
- Cramer P, Bushnell DA, Kornberg RD (2001) Structural basis of transcription: RNA polymerase at 2.8 Å resolution. *Science* **292**: 1863–1876
- Davis CA, Ares Jr M (2006) Accumulation of unstable promoter-associated transcripts upon loss of the nuclear exosome subunit Rrp6p in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci USA* **103**: 3262–3267
- Escobar-Henriques M, Daignan-Fornier B, Collart MA (2003) The critical cis-acting element required for IMD2 feedback regulation by GDP is a TATA box located 202 nucleotides upstream of the transcription start site. *Mol Cell Biol* **23**: 6267–6278
- Exinger F, Lacroute F (1992) 6-Azauracil inhibition of GTP biosynthesis in *Saccharomyces cerevisiae*. *Curr Genet* **22**: 9–11
- Flynn PJ, Reece RJ (1999) Activation of transcription by metabolic intermediates of the pyrimidine biosynthetic pathway. *Mol Cell Biol* **19**: 882–888
- Ghavi-Helm Y, Michaut M, Acker J, Aude J, Thuriaux P, Werner M, Soutourina J (2008) Genome-wide location analysis revealed a role for TFIIS in RNA polymerase III transcription. *Genes Dev* **14**: 1934–1947
- Gnatt AL, Cramer P, Fu J, Bushnell DA, Kornberg RD (2001) Structural basis of transcription: an RNA polymerase elongation complex at 3.3 Å resolution. *Science* **292**: 1876–1882
- Gong XQ, Zhang C, Feig M, Burton ZF (2005) Dynamic error correction and regulation of downstream bubble opening by human RNA polymerase II. *Mol Cell* **18**: 461–470
- Harismendy O, Gendrel CG, Soularue P, Gidrol X, Sentenac A, Werner M, Lefebvre O (2003) Genome-wide location of yeast RNA polymerase III transcription machinery. *EMBO J* **22**: 4738–4747
- Hirata A, Klein BJ, Murakami KS (2008) The X-ray crystal structure of RNA polymerase from Archaea. *Nature* **451**: 851–854
- Hyle JW, Shaw RJ, Reines D (2003) Functional distinctions between IMP dehydrogenase genes in providing mycophenolate resistance and guanine prototrophy to yeast. *J Biol Chem* **278**: 28470–28478
- Jenks MH, O'Rourke TW, Reines D (2008) Properties of an intergenic terminator and start site switch that regulate IMD2 transcription in yeast. *Mol Cell Biol* **28**: 3883–3893
- Juneau K, Palm C, Miranda M, Davis RW (2007) High-density yeast-tiling array reveals previously undiscovered introns and extensive regulation of meiotic splicing. *Proc Natl Acad Sci USA* **104**: 1522–1527
- Kettenberger H, Armache KJ, Cramer P (2003) Architecture of the RNA polymerase II-TFIIS complex and implications for mRNA cleavage. *Cell* **114**: 347–357
- Kopcewicz KA, O'Rourke TW, Reines D (2007) Metabolic regulation of IMD2 transcription and an unusual DNA element that generates short transcripts. *Mol Cell Biol* **27**: 2821–2829
- Lacroute F, Piérard A, Grenson M, Wiame JM (1965) The biosynthesis of carbamoylphosphate in *Saccharomyces cerevisiae*. *J Gen Microbiol* **40**: 127–142
- Landick R (2005) NTP-entry routes in multi-subunit RNA polymerases. *Trends Biochem Sci* **30**: 651–654
- Loison G, Losson R, Lacroute F (1980) Constitutive mutants for orotidine 5'-phosphate decarboxylase and dihydroorotic acid dehydrogenase in *Saccharomyces cerevisiae*. *Curr Genet* **2**: 39–44
- Losson R, Lacroute F (1981) Cloning of a eukaryotic regulatory gene. *Mol Gen Genet* **184**: 394–399
- Miller JH (1972) *Experiments in Molecular Genetics*. Cold Spring Harbor: Cold Spring Harbor Laboratory Press. pp 352–353
- Nadkarni AK, McDonough VM, Yang WL, Stuke JE, Ozier-Kalogeropoulos O, Carman GM (1995) Differential biochemical regulation of the URA7- and URA8-encoded CTP synthetases from *Saccharomyces cerevisiae*. *J Biol Chem* **270**: 24982–24988
- Ozier-Kalogeropoulos O, Adeline MT, Yang WL, Carman GM, Lacroute F (1994) Use of synthetic lethal mutants to clone and characterize a novel CTP synthetase gene in *Saccharomyces cerevisiae*. *Mol Gen Genet* **242**: 431–439
- Potier S, Lacroute F, Hubert JC, Souciet JL (1990) Studies on transcription of the yeast URA2 gene. *FEMS Microbiol Lett* **60**: 215–219
- Potier S, Souciet J-L, Lacroute F (1987) Correlation between restriction map, genetic map and catalytic functions in the gene complex URA2. *Mol Gen Genet* **209**: 283–289
- Rothstein R (1991) Targeting, disruption, replacement, and allele rescue: integrative DNA transformation in yeast. *Methods Enzymol* **194**: 281–301
- Scafe C, Martin C, Nonet M, S P, Okamura S, Young RA (1990) Conditional mutations occur predominantly in highly conserved residues of RNA polymerase II subunits. *Mol Cell Biol* **10**: 1270–1275
- Serre V, Penverne B, Souciet JL, Potier S, Guy H, Evans D, Vicart P, Herve G (2004) Integrated allosteric regulation in the *S. cerevisiae* carbamylphosphate synthetase—aspartate transcarbamylase multifunctional protein. *BMC Biochem* **5**: 6
- Shaw RJ, Wilson JL, Smith KT, Reines D (2001) Regulation of an IMP dehydrogenase gene and its overexpression in drug-sensitive transcription elongation mutants of yeast. *J Biol Chem* **276**: 32905–32916
- Snow R (1966) An enrichment method for auxotrophic yeast mutants using the antibiotic 'nystatin'. *Nature* **211**: 206–207
- Soutourina J, Bordas-Le Floch V, Gendrel G, Flores A, Ducrot C, Dumay-Odelot H, Soularue P, Navarro F, Cairns BR, Lefebvre O, Werner M (2006) Rsc4 connects the chromatin remodeler RSC to RNA polymerases. *Mol Cell Biol* **26**: 4920–4933
- Steinmetz EJ, Warren CL, Kuehner JN, Panbehi B, Ansari AZ, Brow DA (2006) Genome-wide distribution of yeast RNA polymerase II and its control by Sen1 helicase. *Mol Cell* **24**: 735–746
- Thiebaut M, Colin J, Neil H, Jacquier A, Seraphin B, Lacroute F, Libri D (2008) Futile cycle of transcription initiation and termination modulates the response to nucleotide shortage in *S. cerevisiae*. *Mol Cell* (in press)
- Thiebaut M, Kisseleva-Romanova E, Rougemaille M, Boulay J, Libri D (2006) Transcription termination and nuclear degradation of cryptic unstable transcripts: a role for the *nrd1-nab3* pathway in genome surveillance. *Mol Cell* **23**: 853–864
- Vassilyev DG, Vassilyeva MN, Perederina A, Tahirov TH, Artsimovitch I (2007) Structural basis for transcription elongation by bacterial RNA polymerase. *Nature* **448**: 157–162
- Westover KD, Bushnell DA, Kornberg RD (2004) Structural basis of transcription: nucleotide selection by rotation in the RNA polymerase II active center. *Cell* **119**: 481–489
- Woychik NA, Lane WS, Young RA (1991) Yeast RNA polymerase II subunit RPB9 is essential for growth at temperature extremes. *J Biol Chem* **266**: 19053–19055
- Woychik NA, Young RA (1989) RNA polymerase II subunit RPB4 is essential for high and low temperature yeast cell growth. *Mol Cell Biol* **9**: 2854–2859
- Zaros C, Briand JF, Boulard Y, Labarre-Mariotte S, Garcia-Lopez C, Thuriaux P, Navarro F (2007) Functional organisation of the Rpb5 subunit shared by the three yeast RNA polymerases. *Nucleic Acids Res* **35**: 634–647
- Zhang Z, Dietrich FS (2005) Mapping of transcription start sites in *Saccharomyces cerevisiae* using 5' SAGE. *Nucleic Acids Res* **33**: 2838–2851



The EMBO Journal is published by Nature Publishing Group on behalf of European Molecular Biology Organization. This article is licensed under a Creative Commons Attribution-NonCommercial-Share Alike 3.0 Licence. [<http://creativecommons.org/licenses/by-nc-sa/3.0/>]

**Supplementary information S1.** Organisation of the pFL80 *pURA2::LacZ* reporter plasmid compared to the vector of origin pFL39-CII. The sequence of 90 nt linker is provided. pFL83 corresponds to pFL80 with the *LacZ* open reading frame replaced by *HIS3*.



## Plasmids

Name	Yeast genetic elements	Origin
pFL36-CII	<i>CEN6 LEU2</i>	pFL36 (Bonneaud <i>et al</i> , 1991)
pFL38-CII	<i>CEN6 URA3</i>	pFL38 (Bonneaud <i>et al</i> , 1991)
pFL38-RPB1	<i>CEN6 URA3 RPB1</i>	Sub-cloning of a <i>SwaI-PshAI RPB1</i> insert in pFL38-CII
pFL39-CII	<i>CEN6 TRP1</i>	pFL39 (Bonneaud <i>et al</i> , 1991)
pFL80	<i>CEN6 TRP1 pURA2::LacZ</i>	<i>KpnI-NotI</i> insertion of <i>pURA2::LacZ</i> in pFL39-CII
pFL80 H2-BSA	<i>CEN6 TRP1 pURA2-BSA::LacZ</i>	<i>BsrGI-SalI</i> deletion in pFL80.
pFL80-H2	<i>CEN6 TRP1 pURA2::LacZ</i>	pFL80 with a 582 bp long <i>HindIII</i> insert (positions 7679-8251 of <i>TEL1</i> )
pFL80-INTA	<i>CEN6 TRP1 pURA2-intA::LacZ</i>	<i>BsmI-BamHI</i> deletion of the <i>pURA2</i> intron in pFL80
pFL80-KBA	<i>CEN6 TRP1 pURA2-KBA::LacZ</i>	<i>KpnI-BsrGI</i> deletion in pFL80
pFL80-upA	<i>CEN6 TRP1 pURA2-upA::LacZ</i>	Subcloning of the <i>KpnI- BamHI</i> fragment of pFL83-upA plasmids
pFL81	<i>TRP1 pURA2::HIS3</i>	pFL80 without the <i>ClaI CEN6</i> insert
pFL81-up2A	<i>TRP1 pURA2-up2A::HIS3</i>	insertion of the <i>up2A</i> allele in pFL81
pFL82	<i>CEN6 LEU2 pURA2::LacZ</i>	<i>KpnI-NotI</i> insertion of <i>pURA2::LacZ</i> in pFL36-CII
pFL83	<i>CEN6 TRP1 pURA2::HIS3</i>	<i>KpnI-NotI</i> insertion of <i>pURA2::HIS3</i> in pFL39-CII
pFL83-upA	<i>CEN6 TRP1 pURA2-upA::HIS3</i>	Deletions upstream of the <i>SalI</i> site in pFL83

# Yeast strains

Strain	Genotype	Origin/Reference
BMA64-1A	MATa <i>ade2-1 can1-100 leu2-3,112 his3-11,15 trp1Δ ura3-1</i>	(Baudin-Baillieu et al, 1997)
BMA64-URA3	MATa <i>ade2-1 can1-100 leu2-3,112 his3-11,15 trp1Δ</i>	URA3 <sup>+</sup> revertant of BMA64-1A
BY4741	MATa <i>leu2Δ0 met15Δ0 his3Δ1 ura3Δ0</i>	Euroscarf *
D535-4D	MATa <i>leu2 trp1 his3-11,15 lys2-801 RPB3-HA:: KanMX4</i>	FL56 x DY236-7D
D535-9D	MATa <i>leu2 trp1 his3-11,15 lys2-801 RPB3-HA:: KanMX4 rpb1-L1397S</i>	FL56 x DY236-7D
D711-13B	MATα <i>leu2 his3Δ1 met15Δ0 trp1 rpb1-L1397S</i>	BY4741 x FL-M9
D712-10C	MATα <i>trp1 leu2 ura3 his3 lys2 RAD3-TAP::URA3</i>	SC0946 x BY4741
D714-5D	MATα <i>leu2 ura3 his3 met15Δ0 SUA7-TAP::URA3</i>	SC0415 x BY4741
D876-10D	MATa <i>his3 leu2-3,112 trp1 leu2 rpb1-L1397S up2Δ</i>	FL51x FL-M9
DY263-7D	MATa <i>leu2-Δ1 lys2-801 trp1Δ63</i>	FY67 x YPH500
FL50	MATα <i>ade2-1 can1-100 leu2-3,112 his3-11,15 ura3-1 trp1-1::pFL81</i> ( <i>TRP1 pURA2::HIS3</i> )	pFL81 in the <i>trp1-1</i> allele (W303-1B)
FL51	MATa <i>ade2-1 can1-100 leu2-3,112 trp1Δ his3-11,15 up2Δ</i>	this work
FL52	MATa <i>ade2-1 can1-100 leu2-3,112 his3-11 TRP1::pFL81::trp1-1</i> // <i>pFL82 (CEN6 LEU2 pURA2::LacZ)</i>	this work
FL53	MATa <i>ade2-1 can1-100 leu2-3,112 trp1-Δ his3-11 ppr1::KanMX4</i>	this work
FL55	MATa <i>ade2-1 can1-100 leu2-3,112 his3-11,15 trp1Δ ura7::HIS3 ura8ts1</i>	this work
FL56	MATα <i>ade2-1 can1-100 leu2-3,112 his3-11,15 trp1Δ</i> <i>RPB3-HA::KanMX4 rpb1-L1397S</i>	<i>RPB3-HA::KanMX4</i> gene replacement in FL-M9
FL-M10	MATα <i>ade2-1 can1-100 leu2-3,112 trp1-1 his3-11,15 ura3-1 rpb1-S1401P</i>	UV mutagenesis of FL50
FL-M13/M23	MATα <i>ade2-1 can1-100 leu2-3,112 trp1-1 his3-11,15 ura3-1 rpb1-F1402L</i>	UV mutagenesis of FL50
FL-M9	MATα <i>ade2-1 can1-100 leu2-3,112 trp1-1 his3-11,15 ura3-1 rpb1-L1397S</i>	UV mutagenesis of FL50
FL-M9-URA3	MATa <i>ade2-1 can1-100 leu2-3,112 trp1-1 his3-11,15 rpb1-L1397S</i>	FL-M9 x BMA64-URA3
FY67	MATa <i>trp1Δ63</i>	(Winston et al, 1995)
GR44-11C	MATa <i>his3-Δ200 trp1-Δ63 lys2-801</i>	FY67 x YPH500
SC0415	MATa <i>ade2 arg4 leu2-3,112 trp1-289 ura3-52 SUA7-TAP::URA3</i>	Euroscarf*
SC0946	MATa <i>ade2 arg4 leu2-3,112 trp1-289 ura3-52 RAD3-TAP::URA3</i>	Euroscarf*

W303-1B	MAT $\alpha$ <i>ade2-1 can1-100 leu2-3,112 trp1-1 his3-11,15 ura3-1</i>	(Rothstein, 1991)
YMW304-8C	MAT $\alpha$ <i>leu2 lys2-801 trp1-1 his3 RPB3-HA::KanMX4 rpb2-<math>\Delta</math>297::HIS3 / rpb2-P1018S (CEN6 LEU2)</i>	Z106 x D535-4D
YPH500	MAT $\alpha$ <i>ade2-1 lys2-801 ura3-52 trp1 <math>\Delta</math>63 his3 <math>\Delta</math>200 leu2 <math>\Delta</math></i>	(Sikorski and Hieter, 1989)
YGH2	MAT $\alpha$ <i>lys2-801 trp1 <math>\Delta</math>63 his3 <math>\Delta</math>200 leu2 <math>\Delta</math>1 TRP1::pADHI::3HA::DST1</i>	Ghavi-Helm et al, in press
YVV9	MAT $\alpha$ <i>ura3-52 his3-D200 leu2 lys2 ade2-1 rpb9::HIS3 trp1-D63 pCM185-RPB9 (pTET::RPB9 CEN4 TRP1)</i>	(Van Mullem et al, 2002)
YWM305-2D	MAT $\alpha$ <i>his3 leu2 lys2 trp1 rpb9::HIS3 RPB3-HA::KanMX4</i>	D535-4D x YVV9
Z106	MAT $\alpha$ <i>ura3-52 his3-D200 leu2-3,112 rpb2-D297::HIS3 / CEN LEU2 rpb2-10 (rpb2-P1018S)</i>	(Scafe et al, 1990)

Euroscarf: <http://web.uni-frankfurt.de/fb15/mikro/euroscarf/>

- Baudin-Baillieu A, Guillemet E, Cullin C and Lacroute F (1997) Construction of a yeast strain deleted for the TRP1 promoter and coding region that enhances the efficiency of the polymerase chain reaction-disruption method. *Yeast* **13**: 353-356
- Rothstein R (1991) Targeting, disruption, replacement, and allele rescue: integrative DNA transformation in yeast. *Methods Enzymol* **194**: 281-301
- Scafe C, Martin C, Nonet MSP, Okamura S, Young RA (1990) Conditional mutations occur predominantly in highly conserved residues of RNA polymerase II subunits. *Mol Cell Biol* **10**: 1270-1275
- Sikorski RS, Hieter P (1989) A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. *Genetics* **122**: 19-27
- Van Mullem V, Wery M, Werner M, Vandenhoute J, Thuriaux P (2002) The Rpb9 subunit of RNA polymerase II binds TFIIE and functionally interacts with the SAGA and Elongator Factors. *J Biol Chem* **277**: 10220-10225
- Winston F, Dollard C and Ricupercio-Hovasse S (1995) Construction of a set of convenient *Saccharomyces cerevisiae* strains that are isogenic to S288C. *Yeast* **11**: 53-55



## Supplementary information S2.

**Sequence conservation of the *URA2* promoter region.** The region comprised between the upstream (Thiebaut et al, submitted) and upstream initiator (this study, shown in green) of the *URA2* promoter region (*S. cerevisiae*) were compared to the corresponding *S. mikatae*, *S. bayanus*, *S. castellii* and *S. kluyveri* domains ([www.yeastgenome.org/](http://www.yeastgenome.org/)). 0 = CUT 5'-ends, as defined by Thiebaut et al, submitted \* = *URA2* mRNA 5'-ends, determined in this study. TCTT (UCUU) boxes are underlined. The region deleted in *up2Δ* is shaded in grey. Boxes indicate the A, T, G, C content of these two regions in *S. cerevisiae*. Dot matrices compare the *TRK1-URA2* intergenic regions of the same yeast species.

```

          0 0 0      0      0
Scer -692  tataaatatcatatctagtatgttttaact-acacca
Smik -682  taaaatgtcatatccagtatgttttaact-acacca
Sbay -720  tagtatatcatttctagcatgttttagcc-acacca
Scas -569  ttcaaaataaaaagaggtatgtttttatcatacca
Sklu -850  tcgttatctctgctatcatgttttttcta-cttcat

```

```

Scer -657  attttctttcaaattatctctt-----tccta--atgggtcttgatattttactttat--tcttgattgt
Smik -647  attttctttcaaattatcatcttttcata-----acatt---tccctcttgatatttctaactgt--tcttgatcgt
Sbay -685  attttctttcaaattatcatcttt-----tctgttgggttcttgatatttca-tcttggt--tcttgaaatgt
Scas -533  ttttctttctaattacttttgtatattattttctttgtagctcttattatattattttatatttcttcttttcttggtttt
Sklu -813  tatttcttgta-----ttcttgtact-----tcttggtaaaaacctctcaaggcactttctacaaggcatttc

```

```

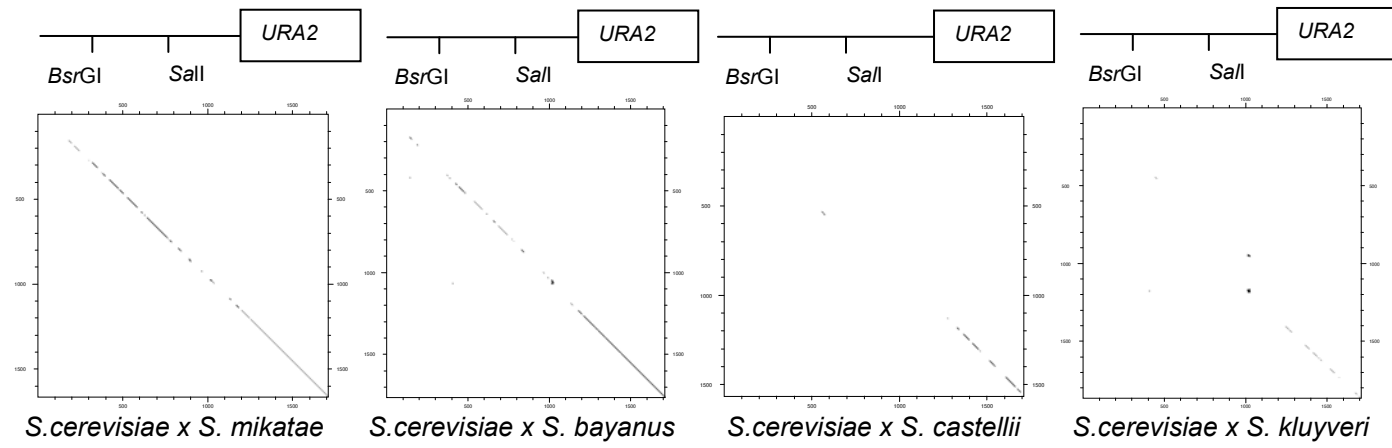
          SalI
          *      * *      *      * *
Scer -596  ag-aaataaaaaacgtGTCGACaacaaccacaaacaaaagg -555
Smik -581  ---gagactaaaaaacGTCGACaacaaccacaaacaaaagg -542
Sbay -623  aaagacttaaaaaaacGTCGACa-caaccacaaacaaaagg -582
Scas -443  aaagggaataag-gacccaacaagc-catcacctcatcgcaac -402
Sklu -749  ccaccactct-agagagaaatcaa-gcaaggaaaacagaccc -709

```

```

A = 24/96 (25%)      A = 24/41 (59%)
T = 49/96 (51%)      T = 03/41 (07%)
G = 08/96 (08%)      G = 05/41 (12%)
C = 15/96 (16%)      C = 09/41 (22%)

```



# Supplementary information S3.

## Primers *URA2/TRK1*

Primer	Sequence (5'-3')	Position (ATG <i>URA2</i> )
TRK1-O1	GACGACGAGGAAAACGAGAG	-1864 to -1844
TRK1-O2	TAAATGGGCACCGATAAAGG	-1734 to -1714
TRK1-O3	ACCTTGAGGGCATGAAATTG	-1395 to -1375
TRK1-O4	TTTCGGGAATGTGTGGTCTT	-1279 to -1259
URA2-A1	CCGGTGTCTCCGATGAAT	-894 to -875
URA2-A2	CGATGAGTGGGGATTTTGTC	-764 to -744
URA2-I1	TCTCTTTCCTTCTCTATTTTGCTCA	-722 to -697
URA2-I2	TGTGGTTGTTGTCGACAGGT	-584 to -564
URA2-A3	ACCTGTCGACAACAACCACA	-584 to -564
URA2-A4	TCAAAGTTGAACGTGGTGAGTC	-454 to -432
URA2-P3	CTGCGATGATGTCGTTTTTC	-257 to -237
URA2-P4	CAAGCCAGATGCCGATATAAA	-126 to -105
URA2-O1	TTCGGTATTTGTCTCGGTCA	895 to 914
URA2-O2	CGAACCCATGGTTTTGAGAT	1023 to 1042
URA2-O3	GCATTTTCATCCTCACCTCAA	1870 to 1890
URA2-O4	TGGAATCACCAGGTGTGAATA	2004 to 2023
URA2-O5	CGAAGTTGCCATCCGTAGAT	2790 to 2809
URA2-O6	CCGTGGTCATCAAAGGATAAA	2922 to 2942
URA2-O7	CCAATTCATTGCAAAGGACA	3744 to 3763
URA2-O8	GATAAGGCGTCAAAGGCAAA	3873 to 3892
URA2-O9	AGTGGCCAACTGAAAAGCAA	4865 to 4884
URA2-O10	TTGACCGTCATGATCAAAGC	4990 to 5009

## Primers *URA1 and URA3*

Primer	Sequence (5'-3')	Position (ATG <i>URA1</i> )
URA1-P1	CCTTGGGATCCAAGTCAATC	-421 to -401
URA1-P2	AGGCTGTAGTGGCTGAAGGA	-273 to -253
URA1-O1	CCTGATGCACCTGCTATTTTC	280 to 300
URA1-O2	GTTTCCCAGGCACATTAGGA	399 to 418
URA1-O3	GTCCGGTAAGGATGCATTTG	765 to 784
URA1-O4	CAACTTCCCACGGAAGTAT	914 to 933

Primer	Sequence (5'-3')	Position (ATG <i>URA3</i> )
URA3-P1	GATTCGGTAATCTCCGAGCA	-156 to -136
URA3-P2	TCTTCGTTTCTGCAGGTTT	-26 to -6
URA3-O1	GGATGTTTCGTACCACCAAGG	108 to 127
URA3-O2	TTGGCGGATAATGCCTTTAG	229 to 248
URA3-O3	GATGACAAGGGAGACGCATT	619 to 638
URA3-O4	CCCAGCCTGCTTTTCTGTAA	750 to 769

## Primers *ADH1*

Primer	Sequence (5'-3')	Position (ATG <i>ADH1</i> )
ADH1-P3	TTCCTTCATTACGCACACT	-232 to -212
ADH1-P4	AGGGAACGAGAACAAATGACG	-91 to -71
ADH1-O1	GGGTATTGACGGTGGTGAAG	597 to 616
ADH1-O2	AAACGTTGATGACACCGTGA	720 to 739
ADH1-O5	GGTGCCAAGTGTTGTCTGA	820 to 839
ADH1-O6	GACAAGCCGACAACCTTGAT	952 to 971
ADH1-T1	CCCCATTTACCCAATTGTA	1264 to 1283
ADH1-T2	ACTGAAGGCTAGGCTGTGGA	1377 to 1396

**Primers *PYK1***

Primer	Sequence (5'-3')	Position (ATG <i>PYK1</i> )
PYK1-P3	CCTTTCCTTCCCATATGATGC	-298 to -277
PYK1-P4	ACTTTGAAAGGGGACCATGA	-144 to -124
PYK1-O5	CACCACCGATGACAAGTACG	333 to 352
PYK1-O6	TTGTCGTCAACGACTTCCAA	463 to 482
PYK1-O1	ATGGTTGCCAGAGGTGACTT	781 to 800
PYK1-O2	TCTGGTTGGTCTTGGGTTGT	917 to 936
PYK1-T1	TCAGGATGGCGAAAGGATAC	1803 to 1822
PYK1-T2	GAAGGCCAGAAGCTGAACAT	1920 to 1939



## **CONCLUSIONS ET PERSPECTIVES**



## Rôle de TFIIS dans la transcription par l'ARN polymérase II

TFIIS est un facteur d'élongation de l'ARN polymérase II, impliqué dans la stimulation du clivage du transcrit lorsque l'ARN polymérase II rencontre des sites d'arrêt de la transcription. Bien que le rôle de ce facteur ait été bien caractérisé *in vitro* au cours de ces trente dernières années, son mode d'action *in vivo* a été très peu étudié. Ainsi, deux hypothèses étaient principalement envisagées quand à son recrutement: TFIIS pourrait être recruté uniquement lorsque l'ARN polymérase II est arrêtée, ou associé et dissocié de façon dynamique au complexe d'élongation de l'ARN polymérase II.

Au cours de ma thèse, nous avons montré par des expériences de localisation à grande échelle de l'ARN polymérase II et du facteur TFIIS, que TFIIS est présent sur près de la moitié des gènes transcrits par l'ARN polymérase II. Par ailleurs, il y a une bonne corrélation entre les valeurs d'enrichissement de TFIIS et de l'ARN polymérase II, bien que dans le cas de TFIIS, la capacité à immunoprécipiter la chromatine de cette protéine soit plus faible. L'analyse des profils d'occupation de TFIIS sur différents gènes de classe II révèle que la protéine est présente tout au long du gène, de façon corrélée à la présence de l'ARN polymérase II. Enfin, la carence en nucléotides induite par la présence de MPA n'a que peu d'effet sur les taux d'enrichissement de TFIIS sur le génome. Par conséquent, nos résultats sont plutôt en faveur du second modèle, selon lequel TFIIS serait recruté de façon dynamique, et permettrait à l'ARN polymérase II de changer de conformation lorsqu'elle rencontre un site d'arrêt. Cependant, la principale limitation à cette conclusion est que nous ignorons la fréquence des pauses ou des arrêts transcriptionnels *in vivo*. En effet, bien que différentes séquences ou conditions entraînant un arrêt de l'ARN polymérase II *in vitro* aient été identifiées (voir Introduction Chapitre II), la position exacte et la fréquence de ces arrêts *in vivo* n'est pas connue. Il est envisageable que, *in vivo*, l'ARN polymérase II soit sujette à des pauses ou des arrêts très fréquents, nécessitant ainsi la présence de TFIIS tout au long de l'élongation de la transcription. La réponse à cette question nécessiterait la mise au point d'une technique permettant de visualiser de façon précise les sites de pause *in vivo*.

Des travaux récents ont mis en évidence un rôle de TFIIS dans l'initiation de la transcription par l'ARN polymérase II (Malagon et al., 2004; Prather et al., 2005; Kim et al., 2007; Guglielmi et al., 2007). Ce rôle est indépendant de l'activité de clivage de TFIIS, mais requiert son domaine de liaison à l'ARN polymérase II. Le modèle actuel suggère cependant que TFIIS ne serait recruté qu'au promoteur de certains gènes. L'analyse de la localisation

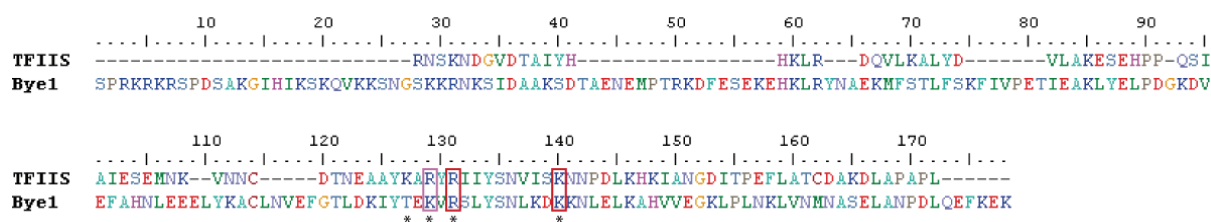


globale de TFIIS sur le génome pourrait révéler sur quelles catégories de promoteurs cette protéine est recrutée. La résolution des puces à ADN utilisées dans notre étude ne permet malheureusement pas de conclure quand à la présence de TFIIS sur les régions promotrices des gènes. En effet, nous ne pouvons distinguer avec certitude la présence de TFIIS sur la région promotrice de sa présence au début de la région codante. Cependant, nous pourrions envisager de mieux comprendre le rôle de TFIIS dans l'initiation de la transcription, et notamment déterminer s'il joue un rôle uniquement pour certains groupes de gènes, en analysant la localisation à grande échelle de TFIIS grâce à des puces "tiling" ou par ChIP-seq.

De nombreuses autres questions restent ouvertes quand au rôle de TFIIS dans l'initiation de la transcription par l'ARN polymérase II. En particulier, nous ignorons si TFIIS interagit avec l'ARN polymérase II de la même façon en initiation et en élongation. Le complexe ARN polymérase II – TFIIS adopte-t-il une structure différente dans le PIC et dans un complexe d'élongation? Pourquoi TFIIS serait recruté sur les promoteurs de certains gènes, et comment agit-il, en conjonction avec le Médiateur, sur la formation du PIC? Existe-t-il un lien entre le rôle de TFIIS dans l'initiation et dans l'élongation? Enfin, le rôle du domaine I de TFIIS reste encore à définir. En effet, bien qu'il ait été montré dans un test double hybride que le domaine I de TFIIS lie spécifiquement les sous-unités Med13 du Médiateur et Spt8 de SAGA (Wery et al., 2004), il ne semble pas être absolument requis pour le rôle de TFIIS ni dans l'initiation ni dans l'élongation de la transcription par l'ARN polymérase II.

#### TFIIS et Bye1 : un rôle antagoniste dans la transcription par l'ARN polymérase II?

TFIIS est une protéine extrêmement bien conservée au cours de l'évolution. Comme nous l'avons vu dans l'introduction, les différents domaines de TFIIS présentent des homologies avec d'autres protéines impliquées dans la transcription (voir Introduction Chapitre II). Ainsi, la protéine Bye1 possède un domaine dit "TFIIS-like", présentant 43% de similarité avec le domaine de liaison à l'ADN de TFIIS chez la drosophile. Cette conservation est moins marquée chez *S. cerevisiae*, mais nous pouvons remarquer que certains des acides aminés impliqués dans l'interaction de TFIIS avec l'ARN polymérase II sont conservés (Figure 32). La protéine Bye1, très peu étudiée chez *S. cerevisiae*, est un suppresseur de la prolyl isomerase Ess1 et inhiberait la transcription par l'ARN polymérase II (Wu et al., 2003). Bye1 est l'homologue de la protéine humaine PHF3, dont un rôle possible dans la formation des



**Figure 32: Alignement de la séquence des domaines de liaison à l'ADN de TFIIS et Bye1 chez *S. cerevisiae*.**

Le domaine II de TFIIS et le domaine "TFIIS-like" de Bye1 sont alignés en utilisant le logiciel Muscle (<http://www.ebi.ac.uk/Tools/muscle/index.html>). Les acides aminés impliqués dans l'interaction entre TFIIS et l'ARN polymérase II sont marqués par des astérisques. Parmi ceux-ci, les acides aminés conservés sont entourés en rouge, et les acides aminés homologues en rose. L'acide aminé R200 de TFIIS, conservé entre les deux protéines est ici numéroté 131.

glioblastomes a pu être mis en évidence (Fischer et al., 2001; Struss et al., 2001; Pallasch et al., 2005). Notre hypothèse est que Bye1 est un compétiteur de TFIIS, opérant comme un régulateur négatif de la transcription, dont l'inactivation provoquerait une dérégulation transcriptionnelle à l'origine des glioblastomes chez l'homme.

Afin de vérifier cette hypothèse, nous avons construit une souche de *S. cerevisiae* délétée pour Bye1 et étiquetée pour TFIIS, ainsi qu'une souche étiquetée pour chacune de ces deux protéines. Des expériences d'immunoprécipitation de la chromatine sont actuellement en cours et permettront de déterminer si Bye1 se lie à la chromatine, au niveau de gènes transcrits par l'ARN polymérase II. Sans impliquer nécessairement un contact direct avec l'ARN polymérase II, ces expériences mettront en évidence la présence de Bye1 sur la totalité ou une partie du transcriptome, et peut-être une occupation mutuellement exclusive par TFIIS et Bye1. En parallèle, des expériences de co-immunoprécipitation nous permettront de mettre à jour une éventuelle interaction entre Bye1 et l'ARN polymérase II. La distribution de Bye1 sur un gène donné permettra par ailleurs d'avoir une première indication sur les rôles possibles de cette protéine au cours de l'initiation ou de l'élongation de la transcription.

Afin de vérifier notre hypothèse selon laquelle Bye1 serait un compétiteur de TFIIS pour sa liaison à l'ARN polymérase II, nous envisageons d'étudier si la délétion de Bye1 entraîne une suppression des phénotypes synthétiques de mutants du domaine II de TFIIS avec d'autres facteurs de transcription, ou si sa surexpression entraîne une aggravation de ces phénotypes. Une mutagenèse ciblée de Bye1 pourra préciser les rôles de ses différents domaines dans ces interactions génétiques.

Enfin, l'étude de l'effet de la délétion de *BYE1* sur le transcriptome ou sur la localisation globale de l'ARN polymérase II et du facteur TFIIS nous permettra de mieux comprendre le rôle de cette protéine, et notamment de déterminer si elle agit comme un régulateur négatif de la transcription par l'ARN polymérase II. Ce travail pourra par la suite être étendu au modèle murin, afin d'étudier plus précisément l'implication de PHF3 dans la formation des glioblastomes.

### Rôle de TFIIS dans la transcription par l'ARN polymérase III

L'étude de la localisation globale du facteur TFIIS a également révélée, de façon surprenante, la présence de TFIIS sur tous les gènes transcrits par l'ARN polymérase III, à l'exception de l'ADNr 5S, suggérant ainsi un rôle jusqu'alors inconnu de cette protéine dans la transcription par l'ARN polymérase III. Des expériences *in vivo* et *in vitro* nous ont permis de montrer que le motif RSADE de TFIIS, responsable de son activité de clivage, est requis pour son rôle dans la transcription par l'ARN polymérase III. Nous avons également montré que, *in vitro*, TFIIS est impliqué dans le choix du site d'initiation de la transcription.

Ces données suggèrent que TFIIS aurait un rôle important dans l'initiation de la transcription par l'ARN polymérase III. Il est intéressant de remarquer que TFIIS a également un rôle dans l'initiation de la transcription par l'ARN polymérase II, mais que ce rôle dépend du domaine II de TFIIS, alors que dans le cas de l'ARN polymérase III, nous avons montré qu'il dépendait de son domaine III, et plus spécifiquement du motif RSADE. Il serait néanmoins intéressant de tester d'autres mutants de TFIIS pour leur effet sur la transcription par l'ARN polymérase III, afin de vérifier si d'autres domaines de TFIIS pourraient être impliqués dans la transcription par l'ARN polymérase III. Des résultats préliminaires révèlent que, parmi quelques mutants du domaine III sélectionnés, seuls des mutants du motif RSADE présentent un phénotype de croissance ralentie à 16°C. Cette condition affectant cependant à la fois la transcription par les ARN polymérases II et III, des expériences complémentaires de ChIP devront être envisagées.

Lors de notre étude, nous n'avons pas pu mettre en évidence, par CoIP, d'interaction directe entre TFIIS et l'un des acteurs de la machinerie de transcription par l'ARN polymérase III. Cependant, des expériences de CoIP après pontage au formaldéhyde révèlent la présence simultanée de TFIIS avec l'ARN polymérase III et le facteur TFIIC. L'hypothèse la plus

probable reste toutefois que TFIIS interagisse directement avec l'ARN polymérase III, mais que cette interaction soit transitoire. Différentes techniques telles que le FRET pourraient être envisagées pour identifier une telle interaction. Par ailleurs, nous pourrions tester l'interaction entre TFIIS et ses partenaires potentiels dans des conditions rendant la présence de TFIIS importante pour la transcription par l'ARN polymérase III, par exemple en réalisant des expériences de CoIP à partir d'extraits de levures cultivées à 16°C. Nous pourrions également mettre en évidence une interaction *in vitro* par "gel mobility shift assay" (Awrey et al., 1998). Cette expérience consiste à incuber la protéine TFIIS purifiée marquée radioactivement avec un de ses partenaires potentiels et analyser la migration du complexe éventuellement formé sur un gel d'acrylamide natif. Par ailleurs, nous pourrions rechercher des mutants de l'ARN polymérase III affectant spécifiquement la présence de TFIIS sur les gènes de classe III, et qui seraient ainsi potentiellement impliqués dans une interaction avec TFIIS. Nous pourrions tester des mutants déjà caractérisés de l'ARN polymérase III, et rechercher des mutations de l'ARN polymérase III entraînant un phénotype de cryosensibilité ou un phénotype synthétique avec des mutants de *DST1*.

Si TFIIS interagit directement avec l'ARN polymérase III, l'obtention de la structure cristalline de l'enzyme serait un pas important dans la compréhension du rôle de TFIIS dans la transcription des gènes de classe III. En effet, la structure de l'ARN polymérase II interagissant avec TFIIS a montré que cette protéine peut ouvrir une crevasse dans l'entonnoir ("funnel") de l'ARN polymérase II, afin d'apporter son motif RSADE à proximité du site actif de l'enzyme. La possibilité de créer une telle crevasse dans la structure de l'ARN polymérase III suggérerait fortement que TFIIS puisse avoir un rôle similaire dans la transcription par l'ARN polymérase III. Une autre possibilité serait que ce soit la sous-unité Rpc11 de l'ARN polymérase III qui adopte une structure proche de celle de TFIIS. Rpc11 est responsable de l'activité de clivage de l'ARN polymérase III, et est également impliquée dans la réinitiation de la transcription par l'ARN polymérase III. Les modèles actuels suggèrent que Rpc11 adopterait une structure homologue à celle de la sous-unité Rpb9 de l'ARN polymérase II. Il paraît cependant assez improbable qu'une protéine impliquée dans le clivage du transcrit ne contacte pas directement le site actif de l'ARN polymérase III. Or la sous-unité Rpb9, bien qu'interagissant avec Rpb1 ne se trouve pas à proximité immédiate du site actif. Si l'extrémité C-terminale de Rpc11 pénètre jusqu'au site actif de l'ARN polymérase III, alors il est possible que TFIIS adopte une position plus périphérique (suggérée par la difficulté à coimmunoprécipiter TFIIS et l'ARN polymérase III) proche de celle de Rpb9 dans la structure de l'ARN polymérase II.

Les résultats obtenus au cours de ma thèse indiquent que TFIIS serait un facteur de transcription de l'ARN polymérase III impliqué dans le choix du site d'initiation. Cette propriété n'a cependant pu être montrée qu'*in vitro*. Des expériences d'extension d'amorce réalisées *in vivo* indiquent que la mutation *dst1-E291A* ne perturbe pas le choix du site d'initiation. La première hypothèse envisagée est que d'autres protéines, absentes lors des expériences réalisées *in vitro*, masqueraient l'effet de cette mutation. Notre premier candidat a été la protéine Nhp6, dont un rôle dans le choix du site d'initiation de la transcription a été décrit *in vitro* et *in vivo* (Kassavetis & Steiner, 2006). Nous avons observé que le triple mutant *dst1Δ nhp6aΔ nhp6bΔ* est viable, et que la délétion de *DST1* n'entraîne pas d'aggravation dans le choix incorrect du site d'initiation par rapport au double mutant. Il est donc probable que d'autres protéines soient impliquées dans le choix du site d'initiation de l'ARN polymérase III. La recherche de gènes dont la délétion concomitante à celle de *DST1* induirait une aggravation dans le choix incorrect du site d'initiation permettrait de découvrir de nouvelles protéines impliquées dans la transcription par l'ARN polymérase III.

Le rôle exact de TFIIS *in vivo* dans la transcription par l'ARN polymérase III reste donc encore à déterminer. Les données de localisation globale de l'ARN polymérase III montrent que la délétion de *DST1* ou la mutation *dst1-E291A* entraînent une diminution de la présence de l'enzyme sur ses gènes cible. TFIIS pourrait ainsi induire un changement de conformation de l'ARN polymérase III la rendant plus stable. Par ailleurs, bien que nous ayons montré un rôle de TFIIS dans le choix du site d'initiation *in vitro*, nous ne pouvons écarter une implication possible de cette protéine au cours d'autres étapes de l'initiation (recrutement de l'ARN polymérase III, formation du PIC, initiation abortive, départ du promoteur...). De plus, il reste envisageable que l'activité de clivage de TFIIS participe à une étape précoce de l'initiation de la transcription par l'ARN polymérase III. La réponse à ces questions pourra être abordée *in vitro*, en décomposant les différentes étapes de l'initiation de la transcription par l'ARN polymérase III. Afin de mieux comprendre les mécanismes mis en jeu *in vivo*, des tests de transcription *in vitro* pourront être réalisés sur des matrices chromatinienne reconstituées.

Notre vision de la machinerie de transcription par l'ARN polymérase III est en constante évolution. En effet, s'il y a quelques années nous pouvions opposer la transcription par l'ARN polymérase II, avec les nombreuses protéines impliquées dans son mécanisme et sa régulation, à la transcription par l'ARN polymérase III, beaucoup plus simple et impliquant peu de facteurs, ce n'est aujourd'hui plus le cas. La transcription par l'ARN polymérase III est sans doute tout aussi complexe que celle par l'ARN polymérase II, mais ayant été moins

étudiée, un nombre important de protéines impliquées dans sa régulation reste probablement encore inconnu. Ainsi, la découverte du rôle de la protéine Maf1 dans la régulation de la transcription par l'ARN polymérase III date d'il y a moins de 10 ans (Pluta et al., 2001). Nous pourrions envisager différentes stratégies de découverte de nouvelles protéines impliquées dans la transcription par l'ARN polymérase III. L'une d'entre elles serait d'étudier la composition de la fraction B'' par fractionnement et spectrométrie de masse. En effet, cette fraction obtenue lors de la purification de la protéine Bdp1 comporte probablement d'autres protéines impliquées dans la transcription par l'ARN polymérase III. Nous avons ainsi montré par western blot que TFIIS était présent dans la fraction B''. Une approche complémentaire serait de développer une matrice immobilisée de transcription par l'ARN polymérase III et de déterminer les protéines présentes dans le complexe de préinitiation par spectrométrie de masse comme cela a été fait pour l'ARN polymérase II (Ranish et al., 2003; Kim et al., 2007).

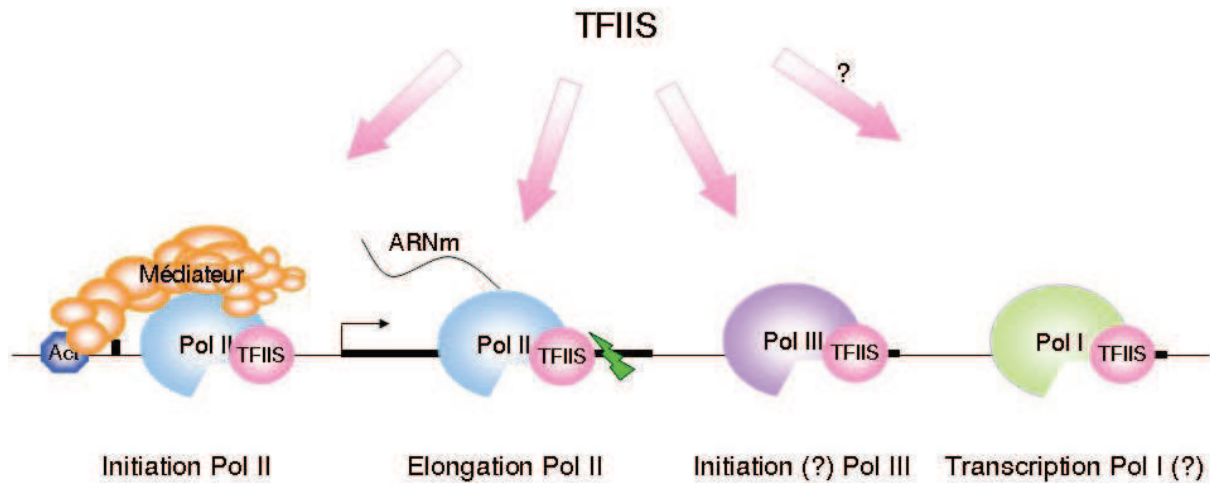
#### TFIIS : un facteur commun aux trois machineries de transcription eucaryotiques?

La protéine TFIIS est connue depuis de nombreuses années comme étant un facteur de transcription de l'ARN polymérase II. Les résultats obtenus au cours de ma thèse ont permis de montrer qu'elle agissait également comme facteur de transcription de l'ARN polymérase III. Nous nous sommes alors interrogés sur un rôle possible de TFIIS dans la transcription par l'ARN polymérase I (Figure 33). Des expériences de ChIP montrent un faible enrichissement de TFIIS sur l'ADNr 35S par rapport aux régions intergéniques NTS1 et NTS2. Par ailleurs, il a été précédemment suggéré que TFIIS pourrait être impliqué dans la transcription par l'ARN polymérase I (Sawadogo et al., 1980, 1981; Schnapp et al., 1996), bien que l'activité de clivage de cette enzyme soit portée par sa sous-unité Rpa12 (Kuhn et al., 2007).

Il serait particulièrement intéressant de vérifier si TFIIS a bien un rôle dans la transcription par l'ARN polymérase I, et si tel est le cas s'il s'agit d'un rôle dans son activité de clivage, dans le choix du site d'initiation ou bien encore un autre rôle dans l'initiation ou l'élongation de la transcription. En parallèle, le rôle de TFIIS dans la transcription de l'ADNr 5S devra être plus amplement étudié. Des expériences de ChIP montrent que TFIIS n'est pas présent sur l'ADNr 5S, suggérant qu'il n'est pas nécessaire à la transcription de ce gène. Cependant, et de façon contradictoire, la transcription de l'ADNr 5S est fortement affectée dans le mutant *dst1-E291A* à 16°C. Il est possible que le facteur TFIIS ne soit pas requis en

conditions standard de croissance pour la transcription de l'ADNr 5S, mais que sa présence devienne nécessaire dans des conditions particulières, comme par exemple à 16°C. L'ADNr 5S étant le seul gène nécessitant l'implication du facteur TFIIA, il est possible que le comportement particulier de ce gène soit dû à cette différence. Si TFIIS interagit directement avec l'ARN polymérase III, il paraît peu probable que le facteur TFIIA empêche cette interaction. Cependant, la mise en jeu de TFIIA dans la transcription de l'ADNr 5S pourrait induire un changement de conformation de l'ARN polymérase III (ou d'un autre facteur avec lequel TFIIS interagirait) rendant son interaction avec TFIIS impossible dans certaines conditions de culture. Nous pourrions tenter de répondre à cette question en étudiant l'effet de TFIIS sur la transcription de l'ADNr 5S *in vitro*.

Il n'est pas très surprenant de découvrir des facteurs de transcription impliqués dans les trois systèmes de transcription nucléaire. En effet, la structure générale des ARN polymérases reste globalement conservée, plusieurs sous-unités étant d'ailleurs communes, et leur cycle de transcription est similaire. TFIIS n'est d'ailleurs pas le premier facteur pour lequel un rôle dans différents systèmes de transcription ait été identifié. L'exemple le plus connu est la TBP, qui est impliquée dans la reconnaissance de la région promotrice lors de la transcription par les ARN polymérases I, II et III (Davidson et al., 2004). Il a également été montré que le facteur TFIIF, en plus de ses rôles dans la transcription par l'ARN polymérase II et la réparation de l'ADN, serait également impliqué dans la transcription par l'ARN polymérase I (Iben et al., 2002; Bradsher et al., 2002; Hoogstraten et al., 2002). De même, les protéines Spt4 et Spt5, des facteurs d'élongation de la transcription par l'ARN polymérase II, seraient également impliquées dans l'élongation de la transcription par l'ARN polymérase I et la maturation des ARN ribosomiaux (Schneider et al., 2006). Plus récemment, la protéine Sub1, un coactivateur de la transcription par l'ARN polymérase II, a été impliquée dans la régulation de la transcription par l'ARN polymérase III (Rosonina et al., 2009). Il apparaît donc clairement que les trois systèmes de transcription ont des liens étroits, et nul doute que dans les années à venir nous découvriront de nouveaux facteurs communs à plusieurs systèmes de transcription.



**Figure 33: Fonctions de TFIIS dans la transcription.**

#### Rôle des différentes isoformes de TFIIS chez la souris

La découverte du rôle de TFIIS dans la transcription par l'ARN polymérase III pose la question de la conservation de cette fonction au cours de l'évolution. L'excellente conservation du motif RSADE de TFIIS suggère que les fonctions qui y sont associées le soient aussi. Afin de vérifier cette hypothèse, nous avons entrepris d'étiqueter les trois isoformes murines de TFIIS dans des cellules souches embryonnaires et d'étudier leur localisation à grande échelle par ChIP-seq. Nous avons pu construire des vecteurs de recombinaisons pour ces trois protéines, et obtenir une lignée de cellules ES étiquetées pour la protéine Tcea1. Des tests d'immunoprécipitation de la chromatine sont actuellement en cours. Les premiers résultats indiquent que les taux d'enrichissements obtenus sont faibles, comme cela est le cas chez *S. cerevisiae*. Nous allons donc devoir optimiser le protocole de ChIP afin de disposer de signaux exploitables.

L'analyse de la distribution de la protéine Tcea1 sur le génome des cellules ES murines, et sa comparaison avec celle des ARN polymérases II et III permettra d'avoir un premier aperçu de la conservation des fonctions de TFIIS chez la souris. Cependant, il serait intéressant de compléter cette étude par l'analyse du rôle des deux autres isoformes, Tcea2 et Tcea3, dont l'expression spécifique dans différents tissus pourrait être due à une fonction différente de ces protéines. Pour ceci, il faudra transformer les différents vecteurs de recombinaison obtenus dans des cellules ES de type AT1, ayant la particularité de permettre l'obtention de lignées de souris.



Par la suite, la fonction de ces protéines pourra être plus amplement caractérisée, notamment en étudiant le phénotype, l'effet sur la transcription, et la localisation globale des ARN polymérases II et III dans des mutants de nos gènes d'intérêt, de leur délétion ou de la diminution de leur expression par ARN interférence. Un lien entre l'expression de TFIIS et la formation de certains types de cancers ayant été mis en évidence, l'étude du rôle de cette protéine permettrait de mieux comprendre comment la régulation de son activité pourrait avoir un rôle bénéfique dans des thérapies anticancéreuses.

Nous pourrions également étudier le rôle de TFIIS au cours de la différenciation cellulaire, et notamment dans la formation des neurones. En effet, la lignée de cellules ES utilisées possède un allèle du gène *Sox1* fusionné avec une séquence GFP-ires-Puromycin phosphotransférase (Aubert et al., 2003). Comme *Sox1* est un marqueur précoce de la différenciation neuronale, la présence de cet allèle permet de sélectionner, en présence de puromycine, les cellules qui sont effectivement engagées dans la différenciation neuronale au cours des protocoles de différenciation des cellules ES.

### Rôle du Médiateur dans la mise en place du PIC

Au cours de ma thèse, nous avons mis en évidence une interaction entre la sous-unité Med11 du Médiateur et la sous-unité Rad3 de TFIIH. Afin d'étudier le rôle de cette interaction, nous avons construit différents mutants ponctuels thermosensibles de Med11, abolissant certaines des interactions de cette protéine avec ses partenaires. L'analyse de l'effet de ces mutants sur la mise en place et la localisation à grande échelle de différents acteurs du complexe de préinitiation nous a permis de dresser un modèle de l'assemblage du PIC. Ce modèle, présentant plusieurs états intermédiaires qui diffèrent du modèle précédemment obtenu *in vitro*, indique entre autres que le Médiateur jouerait un rôle important dans le recrutement de TFIIH et TFIIE.

Une méthodologie similaire peut être appliquée à l'étude de toutes les sous-unités essentielles du Médiateur. Elle consiste à isoler des mutants thermosensibles par mutagenèse aléatoire, et à les caractériser pour leur perte d'interaction avec leurs partenaires connus. L'effet de ces mutations sur la localisation globale de différents acteurs de l'initiation de la transcription ou sur le transcriptome pourra ensuite être mis en évidence, et permettre ainsi de

mieux comprendre le mécanisme d'action du Médiateur *in vivo*. Un tel projet vient d'être initié au laboratoire en collaboration avec le laboratoire de Laurent Kuras (CGM / Gif-sur-Yvette).

De façon inattendue, l'étude de la localisation globale de la sous-unité Med5 du Médiateur et des sous-unités Rad3 et Kin28 de TFIIH a révélé la présence de ces deux complexes sur les régions centromériques des 16 chromosomes de *S. cerevisiae*; suggérant un rôle possible dans la ségrégation des chromosomes. Une étude récente, basée sur l'analyse *in silico* de la combinaison de plusieurs cribles à grande échelle, a permis de prédire le rôle potentiel d'une centaine de gènes dans les processus méiotiques qui incluent la réplication de l'ADN, la recombinaison et la ségrégation des chromosomes (Jordan et al., 2007). Les auteurs ont ensuite pu mettre en évidence, par analyse phénotypique de mutants de délétion, que 11 de ces gènes sont impliqués dans la méiose. Ils ont ainsi identifié que la délétion du gène *SOH1*, codant la sous-unité Med31 du Médiateur, entraîne une sensibilité à l'hydroxyurée et une diminution du taux de conversion génique. Par ailleurs, un certain nombre d'interactions génétiques entre *SOH1* et des gènes impliqués dans la réplication, la réparation et la cohésion des chromatides ont été identifiés. Des travaux plus anciens suggéraient également que le gène *CSE2*, codant la sous-unité Med9 du Médiateur, serait impliqué dans la ségrégation des chromosomes (Xiao et al., 1993; Xiao & Fitzgerald-Hayes, 1995). Ces données semblent ainsi suggérer que le Médiateur pourrait avoir un rôle restant à déterminer dans les processus de méiose et de ségrégation des chromosomes. Le complexe TFIIH est quand à lui impliqué dans la réparation et progression du cycle cellulaire, bien que cette dernière activité ne semble pas conservée chez *S. cerevisiae*.

#### Etude de mutants de l'ARN polymérase II: un outil pour la compréhension des mécanismes de la transcription par l'ARN polymérase II

Au cours de ma thèse, j'ai été impliquée dans l'étude de la régulation de la transcription du gène *URA2*. Nous avons identifié deux classes de mutations activant de façon constitutive la transcription du gène *URA2*. Une première classe correspond à de courtes délétions d'une partie de la région promotrice de ce gène, indiquant que cette région serait impliquée dans la régulation négative de la transcription d'*URA2*. Une étude réalisée en parallèle montre que le gène *URA2* possède un site d'initiation alternatif produisant un court transcrit instable (CUT) régulant la transcription de *URA2* (Thiebaut et al., 2008). Ces données suggèrent que le gène *URA2* serait régulé par un mécanisme d'atténuation

comparable à celui décrit pour le gène *IMD2* (Jenks et al., 2008; Kuehner & Brow, 2008). Une deuxième classe de mutations sont situées dans la boucle "Switch 1" de la sous-unité Rpb1 de l'ARN polymérase II. L'analyse du mutant *rpb1-L1397S* révèle une diminution globale de la présence d'ARN polymérase II sur le génome, mais une activation spécifique de certains gènes impliqués dans la biosynthèse des nucléotides tels *URA2*, *IMD2* ou *URA8*. Cette étude met ainsi en évidence une propriété inattendue de l'ARN polymérase II qui agirait comme un senseur de la disponibilité en nucléosides triphosphates dans la cellule.

Cette étude montre la puissance des analyses à grande échelle pour l'étude de la transcription par l'ARN polymérase II. Ce travail pourrait ainsi être étendu à l'analyse d'autres mutants ponctuels des sous-unités essentielles de l'ARN polymérase II ou de mutants de délétion des sous-unités non essentielles Rpb4 et Rpb9. La mise en évidence de modification des profils d'occupation de certains gènes pourraient ainsi nous éclairer sur un rôle de ces sous-unités dans différentes étapes de la transcription. Par ailleurs, l'analyse des gènes dont le profil d'occupation est modifié pourrait suggérer l'implication de l'ARN polymérase II dans différents processus biologiques, tels celui identifié pour la sous-unité Rpb1. Ainsi, l'étude de la sous-unité Rpb9, qui intéresse particulièrement notre équipe, est actuellement en cours au laboratoire.

## **BIBLIOGRAPHIE**



- Adelman K, Marr MT, Werner J, Saunders A, Ni Z, Andrulis ED, & Lis JT (2005) Efficient release from promoter-proximal stall sites requires transcript cleavage factor TFIIS. *Mol. Cell* 17:103-112
- Akhtar MS, Heidemann M, Tietjen JR, Zhang DW, Chapman RD, Eick D, & Ansari AZ (2009) TFIIH kinase places bivalent marks on the carboxy-terminal domain of RNA polymerase II. *Mol. Cell* 34:387-393
- Akoulitchev S, Chuikov S, & Reinberg D (2000) TFIIH is negatively regulated by cdk8-containing mediator complexes. *Nature* 407:102-106
- Andrau JC & Werner M (2001) B"-associated factor(s) involved in RNA polymerase III preinitiation complex formation and start-site selection. *Eur J Biochem* 268:5167-75
- Andrau J, van de Pasch L, Lijnzaad P, Bijma T, Koerkamp MG, van de Peppel J, Werner M, & Holstege FCP (2006) Genome-wide location of the coactivator mediator: Binding without activation and transient Cdk8 interaction on DNA. *Mol Cell* 22:179-92
- Andrulis ED, Guzmán E, Döring P, Werner J, & Lis JT (2000) High-resolution localization of *Drosophila* Spt5 and Spt6 at heat shock genes in vivo: roles in promoter proximal pausing and transcription elongation. *Genes Dev* 14:2635-2649
- Archambault J, Lacroute F, Ruet A, & Friesen JD (1992) Genetic interaction between transcription elongation factor TFIIS and RNA polymerase II. *Mol Cell Biol* 12:4142-52
- Armache KJ, Kettenberger H, & Cramer P (2003) Architecture of initiation-competent 12-subunit RNA polymerase II. *Proc Natl Acad Sci U S A* 100:6964 - 8
- Armache K, Mitterweger S, Meinhart A, & Cramer P (2005) Structures of complete RNA polymerase II and its subcomplex, Rpb4/7. *J. Biol. Chem* 280:7131-7134
- Asturias FJ, Jiang YW, Myers LC, Gustafsson CM, & Kornberg RD (1999) Conserved structures of mediator and RNA polymerase II holoenzyme. *Science* 283:985-987
- Aubert J, Stavridis MP, Tweedie S, O'Reilly M, Vierlinger K, Li M, Ghazal P, Pratt T, Mason JO, Roy D, & Smith A (2003) Screening for mammalian neural genes via fluorescence-activated cell sorter purification of neural precursors from Sox1-gfp knock-in mice. *Proc. Natl. Acad. Sci. U.S.A* 100 Suppl 1:11836-11841
- Awrey DE, Shimasaki N, Koth C, Weilbaecher R, Olmsted V, Kazanis S, Shan X, Arellano J, Arrowsmith CH, Kane CM, & Edwards AM (1998) Yeast transcript elongation factor (TFIIS), structure and function. II: RNA polymerase binding, transcript cleavage, and read-through. *J Biol Chem* 273:22595-605
- Awrey DE, Weilbaecher RG, Hemming SA, Orlicky SM, Kane CM, & Edwards AM (1997) Transcription elongation through DNA arrest sites. A multistep process involving both RNA polymerase II subunit RPB9 and TFIIS. *J Biol Chem* 272:14747-54

- Barrera LO & Ren B (2006) The transcriptional regulatory code of eukaryotic cells--insights from genome-wide analysis of chromatin organization and transcription factor binding. *Curr Opin Cell Biol* 18:291-8
- Barrientos A (2003) Yeast models of human mitochondrial diseases. *IUBMB Life* 55:83-95
- Barski A, Cuddapah S, Cui K, Roh T, Schones DE, Wang Z, Wei G, Chepelev I, & Zhao K (2007) High-resolution profiling of histone methylations in the human genome. *Cell* 129:823-37
- Bengal E & Aloni Y (1989) A block of transcription elongation by RNA polymerase II at synthetic sites in vitro. *J Biol Chem* 264:9791-8
- Bengal E, Flores O, Krauskopf A, Reinberg D, & Aloni Y (1991) Role of the mammalian transcription factors IIF, IIS, and IIX during elongation by RNA polymerase II. *Mol Cell Biol* 11:1195-206
- Bernstein BE, Liu CL, Humphrey EL, Perlstein EO, & Schreiber SL (2004) Global nucleosome occupancy in yeast. *Genome Biol* 5:R62
- Berroteran RW, Ware DE, & Hampsey M (1994) The sua8 suppressors of *Saccharomyces cerevisiae* encode replacements of conserved residues within the largest subunit of RNA polymerase II and affect transcription start site selection similarly to sua7 (TFIIB) mutations. *Mol. Cell. Biol* 14:226-237
- Bhargava P & Kassavetis GA (1999) Abortive initiation by *Saccharomyces cerevisiae* RNA polymerase III. *J. Biol. Chem* 274:26550-26556
- Bhaumik SR, Raha T, Aiello DP, & Green MR (2004) In vivo target of a transcriptional activator revealed by fluorescence resonance energy transfer. *Genes Dev* 18:333-343
- Björklund S & Gustafsson CM (2005) The yeast Mediator complex and its regulation. *Trends Biochem. Sci* 30:240-244
- Bobkova EV, Habib N, Alexander G, & Hall BD (1999) Mutational analysis of the hydrolytic activity of yeast RNA polymerase III. *J. Biol. Chem* 274:21342-21348
- Bondarenko VA, Steele LM, Ujvári A, Gaykalova DA, Kulaeva OI, Polikanov YS, Luse DS, & Studitsky VM (2006) Nucleosomes can form a polar barrier to transcript elongation by RNA polymerase II. *Mol Cell* 24:469-79
- Booth V, Koth CM, Edwards AM, & Arrowsmith CH (2000) Structure of a conserved domain common to the transcription factors TFIIS, elongin A, and CRSP70. *J Biol Chem* 275:31266-8
- Borukhov S, Sagitov V, & Goldfarb A (1993) Transcript cleavage factors from *E. coli*. *Cell* 72:459-66
- Bourbon H (2008) Comparative genomics supports a deep evolutionary origin for the large, four-module transcriptional mediator complex. *Nucleic Acids Res* 36:3993-4008

Bourbon H et al. (2004) A unified nomenclature for protein subunits of mediator complexes linking transcriptional regulators to RNA polymerase II. *Mol. Cell* 14:553-557

Bradsher J, Auriol J, Proietti de Santis L, Iben S, Vonesch JL, Grummt I, & Egly JM (2002) CSB is a component of RNA pol I transcription. *Mol. Cell* 10:819-829

Buck MJ & Lieb JD (2004) ChIP-chip: considerations for the design, analysis, and application of genome-wide chromatin immunoprecipitation experiments. *Genomics* 83:349-360

Bulyk ML (2006) DNA microarray technologies for measuring protein–DNA interactions. *Current Opinion in Biotechnology* 17:422-430

Buratowski S, Hahn S, Guarente L, & Sharp PA (1989) Five intermediate complexes in transcription initiation by RNA polymerase II. *Cell* 56:549-561

Buratowski S (2005) Connections between mRNA 3' end processing and transcription termination. *Curr. Opin. Cell Biol* 17:257-261

Bushnell DA & Kornberg RD (2003) Complete, 12-subunit RNA polymerase II at 4.1-Å resolution: implications for the initiation of transcription. *Proc. Natl. Acad. Sci. U.S.A* 100:6969-6973

Cai G, Imasaki T, Takagi Y, & Asturias FJ (2009) Mediator Structural Conservation and Implications for the Regulation Mechanism. *Structure* 17:559-567

Camblong J, Iglesias N, Fickentscher C, Dieppois G, & Stutz F (2007) Antisense RNA stabilization induces transcriptional gene silencing via histone deacetylation in *S. cerevisiae*. *Cell* 131:706-17

Chadick JZ & Asturias FJ (2005) Structure of eukaryotic Mediator complexes. *Trends Biochem. Sci* 30:264-271

Chang WH & Kornberg RD (2000) Electron crystal structure of the transcription factor and DNA repair complex, core TFIIH. *Cell* 102:609-613

Chapman RD, Heidemann M, Albert TK, Mailhammer R, Flatley A, Meisterernst M, Kremmer E, & Eick D (2007) Transcribing RNA polymerase II is phosphorylated at CTD residue serine-7. *Science* 318:1780-1782

Chapman RD, Heidemann M, Hintermair C, & Eick D (2008) Molecular evolution of the RNA polymerase II CTD. *Trends Genet* 24:289-296

Chédin S, Riva M, Schultz P, Sentenac A, & Carles C (1998) The RNA cleavage activity of RNA polymerase III is mediated by an essential TFIIIS-like subunit and is important for transcription termination. *Genes Dev* 12:3857-71



- Chen JL, Attardi LD, Verrijzer CP, Yokomori K, & Tjian R (1994) Assembly of recombinant TFIID reveals differential coactivator requirements for distinct transcriptional activators. *Cell* 79:93-105
- Cho EJ, Kobor MS, Kim M, Greenblatt J, & Buratowski S (2001) Opposing effects of Ctk1 kinase and Fcp1 phosphatase at Ser 2 of the RNA polymerase II C-terminal domain. *Genes Dev* 15:3319-3329
- Cho EJ, Takagi T, Moore CR, & Buratowski S (1997) mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain. *Genes Dev* 11:3319-3326
- Cho RJ, Campbell MJ, Winzeler EA, Steinmetz L, Conway A, Wodicka L, Wolfsberg TG, Gabrielian AE, Landsman D, Lockhart DJ, & Davis RW (1998) A genome-wide transcriptional analysis of the mitotic cell cycle. *Mol Cell* 2:65-73
- Chu S, DeRisi J, Eisen M, Mulholland J, Botstein D, Brown PO, & Herskowitz I (1998) The transcriptional program of sporulation in budding yeast. *Science* 282:699-705
- Cieřła M & Boguta M (2008) Regulation of RNA polymerase III transcription by Maf1 protein. *Acta Biochim. Pol* 55:215-225
- Clapier CR & Cairns BR (2009) The Biology of Chromatin Remodeling Complexes. *Annu. Rev. Biochem* Available at: <http://www.ncbi.nlm.nih.gov/pubmed/19355820> [Accédé Mai 26, 2009].
- Clark AB, Dykstra CC, & Sugino A (1991) Isolation, DNA sequence, and regulation of a *Saccharomyces cerevisiae* gene that encodes DNA strand transfer protein alpha. *Mol Cell Biol* 11:2576-82
- Collins SR et al. (2007) Functional dissection of protein complexes involved in yeast chromosome biology using a genetic interaction map. *Nature* 446:806-810
- Conaway RC & Conaway JW (1989) An RNA polymerase II transcription factor has an associated DNA-dependent ATPase (dATPase) activity strongly stimulated by the TATA region of promoters. *Proc. Natl. Acad. Sci. U.S.A* 86:7356-7360
- Conaway RC, Sato S, Tomomori-Sato C, Yao T, & Conaway JW (2005) The mammalian Mediator complex and its role in transcriptional regulation. *Trends Biochem. Sci* 30:250-255
- Core LJ, Waterfall JJ, & Lis JT (2008) Nascent RNA Sequencing Reveals Widespread Pausing and Divergent Initiation at Human Promoters. *Science* 322:1845-1848
- Craig AG, Nizetic D, Hoheisel JD, Zehetner G, & Lehrach H (1990) Ordering of cosmid clones covering the herpes simplex virus type I (HSV-I) genome: a test case for fingerprinting by hybridisation. *Nucleic Acids Res* 18:2653-60
- Craighead JL, Chang W, & Asturias FJ (2002) Structure of yeast RNA polymerase II in solution: implications for enzyme regulation and interaction with promoter DNA. *Structure* 10:1117-1125

Cramer P, Bushnell DA, Fu J, Gnatt AL, Maier-Davis B, Thompson NE, Burgess RR, Edwards AM, David PR, & Kornberg RD (2000) Architecture of RNA polymerase II and implications for the transcription mechanism. *Science* 288:640-649

Cramer P, Bushnell DA, & Kornberg RD (2001) Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science* 292:1863 - 76

Cramer P (2004) RNA polymerase II structure: from core to functional complexes. *Curr Opin Genet Dev* 14:218-26

Davidson I, Martianov I, & Viville S (2004) [TBP, a universal transcription factor?]. *Med Sci (Paris)* 20:575-579

Davie JK & Kane CM (2000) Genetic interactions between TFIIS and the Swi-Snf chromatin-remodeling complex. *Mol Cell Biol* 20:5960-73

Davis JA, Takagi Y, Kornberg RD, & Asturias FJ (2002) Structure of the Yeast RNA Polymerase II Holoenzyme: : Mediator Conformation and Polymerase Interaction. *Molecular Cell* 10:409-415

DeRisi JL, Iyer VR, & Brown PO (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* 278:680-6

Dieci G & Sentenac A (1996) Facilitated recycling pathway for RNA polymerase III. *Cell* 84:245-252

Dieci G, Fiorino G, Castelnovo M, Teichmann M, & Pagano A (2007) The expanding RNA polymerase III transcriptome. *Trends in Genetics* 23:614-622

Dieci G & Sentenac A (2003) Detours and shortcuts to transcription reinitiation. *Trends Biochem. Sci* 28:202-209

Ducrot C, Lefebvre O, Landrieux E, Guirouilh-Barbat J, Sentenac A, & Acker J (2006) Reconstitution of the yeast RNA polymerase III transcription system with all recombinant factors. *J. Biol. Chem* 281:11685-11692

Edwards AM, Kane CM, Young RA, & Kornberg RD (1991) Two dissociable subunits of yeast RNA polymerase II stimulate the initiation of transcription at a promoter in vitro. *J. Biol. Chem* 266:71-75

Egloff S, O'Reilly D, Chapman RD, Taylor A, Tanzhaus K, Pitts L, Eick D, & Murphy S (2007) Serine-7 of the RNA polymerase II CTD is specifically required for snRNA gene expression. *Science* 318:1777-1779

Eisen MB, Spellman PT, Brown PO, & Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A* 95:14863-8

- Esnault C, Ghavi-Helm Y, Brun S, Soutourina J, Van Berkum N, Boschiero C, Holstege F, & Werner M (2008) Mediator-dependent recruitment of TFIIF modules in preinitiation complex. *Mol. Cell* 31:337-346
- Exinger F & Lacroute F (1992) 6-Azauracil inhibition of GTP biosynthesis in *Saccharomyces cerevisiae*. *Curr Genet* 22:9-11
- Fabrega C, Shen V, Shuman S, & Lima CD (2003) Structure of an mRNA capping enzyme bound to the phosphorylated carboxy-terminal domain of RNA polymerase II. *Mol. Cell* 11:1549-1561
- Fan X, Chou DM, & Struhl K (2006) Activator-specific recruitment of Mediator in vivo. *Nat Struct Mol Biol* 13:117-20
- Fan X & Struhl K (2009) Where does mediator bind in vivo? *PLoS ONE* 4:e5029
- Feaver WJ, Gileadi O, Li Y, & Kornberg RD (1991) CTD kinase associated with yeast RNA polymerase II initiation factor b. *Cell* 67:1223-1230
- Fernández-Tornero C, Böttcher B, Riva M, Carles C, Steuerwald U, Ruigrok RWH, Sentenac A, Müller CW, & Schoehn G (2007) Insights into transcription initiation and termination from the electron microscopy structure of yeast RNA polymerase III. *Mol. Cell* 25:813-823
- Fischer U, Struss AK, Hemmer D, Michel A, Henn W, Steudel WI, & Meese E (2001) PHF3 expression is frequently reduced in glioma. *Cytogenet. Cell Genet* 94:131-136
- Fish RN, Ammerman ML, Davie JK, Lu BF, Pham C, Howe L, Ponticelli AS, & Kane CM (2006) Genetic interactions between TFIIF and TFIIS. *Genetics* 173:1871-1884
- Fish RN & Kane CM (2002) Promoting elongation with transcript cleavage stimulatory factors. *Biochim Biophys Acta* 1577:287-307
- Fisher RP (2005) Secrets of a double agent: CDK7 in cell-cycle control and transcription. *J. Cell. Sci* 118:5171-5180
- Flanagan PM, Kelleher RJ, Sayre MH, Tschochner H, & Kornberg RD (1991) A mediator required for activation of RNA polymerase II transcription in vitro. *Nature* 350:436-438
- Flores A, Briand JF, Gadai O, Andrau JC, Rubbi L, Van Mullem V, Boschiero C, Goussot M, Marck C, Carles C, Thuriaux P, Sentenac A, & Werner M (1999) A protein-protein interaction map of yeast RNA polymerase III. *Proc. Natl. Acad. Sci. U.S.A* 96:7815-7820
- Fodor SP, Read JL, Pirrung MC, Stryer L, Lu AT, & Solas D (1991) Light-directed, spatially addressable parallel chemical synthesis. *Science* 251:767-73
- Galburt EA, Grill SW, Wiedmann A, Lubkowska L, Choy J, Nogales E, Kashlev M, & Bustamante C (2007) Backtracking determines the force sensitivity of RNAP II in a factor-dependent manner. *Nature* 446:820-3

Gasch AP, Spellman PT, Kao CM, Carmel-Harel O, Eisen MB, Storz G, Botstein D, & Brown PO (2000) Genomic Expression Programs in the Response of Yeast Cells to Environmental Changes. *Mol Biol Cell* 11:4241–4257

Gavin A et al. (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* 415:141-7

Geiduschek EP & Kassavetis GA (2001) The RNA polymerase III transcription apparatus. *J Mol Biol* 310:1-26

Gentleman RC et al. (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* 5:R80

Ghaemmaghami S, Huh W, Bower K, Howson RW, Belle A, Dephoure N, O'Shea EK, & Weissman JS (2003) Global analysis of protein expression in yeast. *Nature* 425:737-41

Giaever G et al. (2002) Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* 418:387-91

Gilmour DS & Lis JT (1985) In vivo interactions of RNA polymerase II with genes of *Drosophila melanogaster*. *Mol Cell Biol* 5:2009-18

Gilmour DS & Lis JT (1986) RNA polymerase II interacts with the promoter region of the noninduced hsp70 gene in *Drosophila melanogaster* cells. *Mol Cell Biol* 6:3984-9

Gilmour DS (2009) Promoter proximal pausing on genes in metazoans. *Chromosoma* 118:1-10

Gnatt AL, Cramer P, Fu J, Bushnell DA, & Kornberg RD (2001) Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Å resolution. *Science* 292:1876-1882

Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M, Louis EJ, Mewes HW, Murakami Y, Philippsen P, Tettelin H, & Oliver SG (1996) Life with 6000 genes. *Science* 274:546, 563-7

Gong XQ, Zhang C, Feig M, & Burton ZF (2005) Dynamic error correction and regulation of downstream bubble opening by human RNA polymerase II. *Mol. Cell* 18:461-470

Grant PA, Duggan L, Côté J, Roberts SM, Brownell JE, Candau R, Ohba R, Owen-Hughes T, Allis CD, Winston F, Berger SL, & Workman JL (1997) Yeast Gcn5 functions in two multisubunit complexes to acetylate nucleosomal histones: characterization of an Ada complex and the SAGA (Spt/Ada) complex. *Genes Dev* 11:1640-1650

Grasser M, Kane CM, Merkle T, Melzer M, Emmersen J, & Grasser KD (2009) Transcript elongation factor TFIIS is involved in arabidopsis seed dormancy. *J Mol Biol* 386:598-611

Grünenfelder B & Winzeler EA (2002) Treasures and traps in genome-wide data sets: case examples from yeast. *Nat Rev Genet* 3:653-61

Gu W, Powell W, Mote J, & Reines D (1993) Nascent RNA cleavage by arrested RNA polymerase II does not require upstream translocation of the elongation complex on DNA. *J Biol Chem* 268:25604-16

Gu W & Reines D (1995) Variation in the size of nascent RNA cleavage products as a function of transcript length and elongation competence. *J Biol Chem* 270:30441-7

Guenther MG, Levine SS, Boyer LA, Jaenisch R, & Young RA (2007) A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* 130:77-88

Guermah M, Kim J, & Roeder RG (2009) Transcription of in vitro assembled chromatin templates in a highly purified RNA polymerase II system. *Methods* Available at: <http://www.ncbi.nlm.nih.gov/pubmed/19272450> [Accédé Avril 27, 2009].

Guermah M, Palhan VB, Tackett AJ, Chait BT, & Roeder RG (2006) Synergistic functions of SII and p300 in productive activator-dependent transcription of chromatin templates. *Cell* 125:275-286

Guglielmi B, van Berkum NL, Klapholz B, Bijma T, Boube M, Boschiero C, Bourbon H, Holstege FCP, & Werner M (2004) A high resolution protein interaction map of the yeast Mediator complex. *Nucleic Acids Res* 32:5379-5391

Guglielmi B, Soutourina J, Esnault C, & Werner M (2007) TFIIS elongation factor and Mediator act in conjunction during transcription initiation in vivo. *Proc Natl Acad Sci U S A* 104:16062-7

Hagler J & Shuman S (1993) Nascent RNA cleavage by purified ternary complexes of vaccinia RNA polymerase. *J. Biol. Chem* 268:2166-2173

Hahn S (2004) Structure and mechanism of the RNA polymerase II transcription machinery. *Nat. Struct. Mol. Biol* 11:394-403

Hallberg M, Polozkov GV, Hu G, Beve J, Gustafsson CM, Ronne H, & Björklund S (2004) Site-specific Srb10-dependent phosphorylation of the yeast Mediator subunit Med2 regulates gene expression from the 2-microm plasmid. *Proc. Natl. Acad. Sci. U.S.A* 101:3370-3375

Hampsey M (1998) Molecular Genetics of the RNA Polymerase II General Transcriptional Machinery. *Microbiol. Mol. Biol. Rev.* 62:465-503

Hanlon SE & Lieb JD (2004) Progress and challenges in profiling the dynamics of chromatin and transcription factor binding with DNA microarrays. *Curr Opin Genet Dev* 14:697-705

Harismendy O, Gendrel C, Soularue P, Gidrol X, Sentenac A, Werner M, & Lefebvre O (2003) Genome-wide location of yeast RNA polymerase III transcription machinery. *EMBO J* 22:4738-47

Hausner W, Lange U, & Musfeldt M (2000) Transcription factor S, a cleavage induction factor of the archaeal RNA polymerase. *J Biol Chem* 275:12393-9

- He Y, Vogelstein B, Velculescu VE, Papadopoulos N, & Kinzler KW (2008) The Antisense Transcriptomes of Human Cells. *Science* 322:1855-1857
- Hekmatpanah DS & Young RA (1991) Mutations in a conserved region of RNA polymerase II influence the accuracy of mRNA start site selection. *Mol. Cell. Biol* 11:5781-5791
- Hemming SA, Jansma DB, Macgregor PF, Goryachev A, Friesen JD, & Edwards AM (2000) RNA polymerase II subunit Rpb9 regulates transcription elongation in vivo. *J. Biol. Chem* 275:35506-35511
- Hengartner CJ, Myer VE, Liao SM, Wilson CJ, Koh SS, & Young RA (1998) Temporal regulation of RNA polymerase II by Srb10 and Kin28 cyclin-dependent kinases. *Mol. Cell* 2:43-53
- Hermann-Le Denmat S, Werner M, Sentenac A, & Thuriaux P (1994) Suppression of yeast RNA polymerase III mutations by FHL1, a gene coding for a fork head protein involved in rRNA processing. *Mol Cell Biol* 14:2905-13
- Ho Y et al. (2002) Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* 415:180-3
- Hoeppner S, Baumli S, & Cramer P (2005) Structure of the mediator subunit cyclin C and its implications for CDK8 function. *J. Mol. Biol* 350:833-842
- Holstege FCP, Jennings EG, Wyrick JJ, Lee TI, Hengartner CJ, Green MR, Golub TR, L ES, er, & Young RA (1998) Dissecting the Regulatory Circuitry of a Eukaryotic Genome. *Cell* 95:717-728
- Hoogstraten D, Nigg AL, Heath H, Mullenders LHF, van Driel R, Hoeijmakers JHJ, Vermeulen W, & Houtsmuller AB (2002) Rapid switching of TFIIH between RNA polymerase I and II transcription and DNA repair in vivo. *Mol. Cell* 10:1163-1174
- Hsu LM (2009) Monitoring abortive initiation. *Methods* 47:25-36
- Hubbard K, Catalano J, Puri RK, & Gnatt A (2008) Knockdown of TFIIS by RNA silencing inhibits cancer cell proliferation and induces apoptosis. *BMC Cancer* 8:133
- Hubert JC, Guyonvarch A, Kammerer B, Exinger F, Liljelund P, & Lacroute F (1983) Complete sequence of a eukaryotic regulatory gene. *EMBO J* 2:2071-3
- Huertas P & Aguilera A (2003) Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. *Mol. Cell* 12:711-721
- Huh W, Falvo JV, Gerke LC, Carroll AS, Howson RW, Weissman JS, & O'Shea EK (2003) Global analysis of protein localization in budding yeast. *Nature* 425:686-91
- Hull MW, McKune K, & Woychik NA (1995) RNA polymerase II subunit RPB9 is required for accurate start site selection. *Genes Dev* 9:481-490

- Iben S, Tschochner H, Bier M, Hoogstraten D, Hozák P, Egly JM, & Grummt I (2002) TFIIF plays an essential role in RNA polymerase I transcription. *Cell* 109:297-306
- Ito T, Chiba T, Ozawa R, Yoshida M, Hattori M, & Sakaki Y (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc Natl Acad Sci U S A* 98:4569-74
- Ito T, Seldin MF, Taketo MM, Kubo T, & Natori S (2000) Gene structure and chromosome mapping of mouse transcription elongation factor S-II (Tcea1). *Gene* 244:55-63
- Ito T, Xu Q, Takeuchi H, Kubo T, & Natori S (1996) Spermatocyte-specific expression of the gene for mouse testis-specific transcription elongation factor S-II. *FEBS Lett* 385:21-24
- Ito T, Arimitsu N, Takeuchi M, Kawamura N, Nagata M, Saso K, Akimitsu N, Hamamoto H, Natori S, Miyajima A, & Sekimizu K (2006) Transcription elongation factor S-II is required for definitive hematopoiesis. *Mol Cell Biol* 26:3194-203
- Iyer VR, Horak CE, Scafe CS, Botstein D, Snyder M, & Brown PO (2001) Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. *Nature* 409:533-8
- Izban MG & Luse DS (1992) The RNA polymerase II ternary complex cleaves the nascent transcript in a 3'----5' direction in the presence of elongation factor SII. *Genes Dev* 6:1342-56
- Izban MG & Luse DS (1993) SII-facilitated transcript cleavage in RNA polymerase II complexes stalled early after initiation occurs in primarily dinucleotide increments. *J Biol Chem* 268:12864-73
- Izban MG & Luse DS (1992) Factor-stimulated RNA polymerase II transcribes at physiological elongation rates on naked DNA but very poorly on chromatin templates. *J Biol Chem* 267:13647-55
- Jasiak AJ, Armache K, Martens B, Jansen R, & Cramer P (2006) Structural biology of RNA polymerase III: subcomplex C17/25 X-ray structure and 11 subunit enzyme model. *Mol. Cell* 23:71-81
- Jenks MH, O'Rourke TW, & Reines D (2008) Properties of an intergenic terminator and start site switch that regulate IMD2 transcription in yeast. *Mol. Cell. Biol* 28:3883-3893
- Jenuwein T & Allis CD (2001) Translating the histone code. *Science* 293:1074-80
- Jeon C & Agarwal K (1996) Fidelity of RNA polymerase II transcription controlled by elongation factor TFIIS. *Proc Natl Acad Sci U S A* 93:13677-82
- Jiang C & Pugh BF (2009) Nucleosome positioning and gene regulation: advances through genomics. *Nat Rev Genet* 10:161-72
- Jordan PW, Klein F, & Leach DRF (2007) Novel roles for selected genes in meiotic DNA processing. *PLoS Genet* 3:e222

Juven-Gershon T, Hsu J, Theisen JW, & Kadonaga JT (2008) The RNA polymerase II core promoter - the gateway to transcription. *Curr. Opin. Cell Biol* 20:253-259

Kash SF, Innis JW, Jackson AU, & Kellems RE (1993) Functional analysis of a stable transcription arrest site in the first intron of the murine adenosine deaminase gene. *Mol Cell Biol* 13:2718-29

Kassavetis GA, Blanco JA, Johnson TE, & Geiduschek EP (1992) Formation of open and elongating transcription complexes by RNA polymerase III. *J. Mol. Biol* 226:47-58

Kassavetis GA, Nguyen ST, Kobayashi R, Kumar A, Geiduschek EP, & Pisano M (1995) Cloning, expression, and function of TFC5, the gene encoding the B" component of the *Saccharomyces cerevisiae* RNA polymerase III transcription factor TFIIB. *Proc. Natl. Acad. Sci. U.S.A* 92:9786-9790

Kassavetis GA & Steiner DF (2006) Nhp6 is a transcriptional initiation fidelity factor for RNA polymerase III transcription in vitro and in vivo. *J Biol Chem* 281:7445-51

Kelleher RJ, Flanagan PM, & Kornberg RD (1990) A novel mediator between activator proteins and the RNA polymerase II transcription apparatus. *Cell* 61:1209-1215

Kellis M, Birren BW, & Lander ES (2004) Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* 428:617-624

Kerppola TK & Kane CM (1990) Analysis of the signals for transcription termination by purified RNA polymerase II. *Biochemistry* 29:269-78

Kettenberger H, Armache K, & Cramer P (2004) Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS. *Mol Cell* 16:955-65

Kettenberger H, Armache K, & Cramer P (2003) Architecture of the RNA Polymerase II-TFIIS Complex and Implications for mRNA Cleavage. *Cell* 114:347-357

Kim B, Nesvizhskii AI, Rani PG, Hahn S, Aebersold R, & Ranish JA (2007) The transcription elongation factor TFIIS is a component of RNA polymerase II preinitiation complexes. *Proc Natl Acad Sci U S A* 104:16068-73

Kim M, Krogan NJ, Vasiljeva L, Rando OJ, Nedeja E, Greenblatt JF, & Buratowski S (2004) The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. *Nature* 432:517-522

Kim TH, Barrera LO, Zheng M, Qu C, Singer MA, Richmond TA, Wu Y, Green RD, & Ren B (2005) A high-resolution map of active promoters in the human genome. *Nature* 436:876-880

Kim YJ, Björklund S, Li Y, Sayre MH, & Kornberg RD (1994) A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. *Cell* 77:599-608



- Kireeva ML, Hancock B, Cremona GH, Walter W, Studitsky VM, & Kashlev M (2005) Nature of the nucleosomal barrier to RNA polymerase II. *Mol Cell* 18:97-108
- Klein C & Struhl K (1994) Increased recruitment of TATA-binding protein to the promoter by transcriptional activation domains in vivo. *Science* 266:280-282
- Komarnitsky P, Cho EJ, & Buratowski S (2000) Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes Dev* 14:2452-2460
- Kornberg RD (2005) Mediator and the mechanism of transcriptional activation. *Trends Biochem. Sci* 30:235-239
- Koschubs T, Seizl M, Larivière L, Kurth F, Baumli S, Martin DE, & Cramer P (2009) Identification, structure, and functional requirement of the Mediator submodule Med7N/31. *EMBO J* 28:69-80
- Koulich D, Orlova M, Malhotra A, Sali A, Darst SA, & Borukhov S (1997) Domain organization of Escherichia coli transcript cleavage factors GreA and GreB. *J. Biol. Chem* 272:7201-7210
- Koyama H, Ito T, Nakanishi T, & Sekimizu K (2007) Stimulation of RNA polymerase II transcript cleavage activity contributes to maintain transcriptional fidelity in yeast. *Genes Cells* 12:547-59
- Kuehner JN & Brow DA (2008) Regulation of a eukaryotic gene by GTP-dependent start site selection and transcription attenuation. *Mol. Cell* 31:201-211
- Kuhn C, Geiger SR, Baumli S, Gartmann M, Gerber J, Jennebach S, Mielke T, Tschochner H, Beckmann R, & Cramer P (2007) Functional architecture of RNA polymerase I. *Cell* 131:1260-72
- Kulish D & Struhl K (2001) TFIIS enhances transcriptional elongation through an artificial arrest site in vivo. *Mol Cell Biol* 21:4162-8
- Kumar A & Snyder M (2001) Emerging technologies in yeast genomics. *Nat Rev Genet* 2:302-12
- Kurtz J, Dufour P, Duclos B, Bergerat J, & Exinger F (2004) [*Saccharomyces cerevisiae*: an efficient tool and model system for anticancer research]. *Bull Cancer* 91:133-9
- Kwapisz M, Wery M, Després D, Ghavi-Helm Y, Soutourina J, Thuriaux P, & Lacroute F (2008) Mutations of RNA polymerase II activate key genes of the nucleoside triphosphate biosynthetic pathways. *EMBO J* 27:2411-2421
- Labhart P & Morgan GT (1998) Identification of novel genes encoding transcription elongation factor TFIIS (TCEA) in vertebrates: conservation of three distinct TFIIS isoforms in frog, mouse, and human. *Genomics* 52:278-88

- Landrieux E, Alic N, Ducrot C, Acker J, Riva M, & Carles C (2006) A subcomplex of RNA polymerase III subunits involved in transcription termination and reinitiation. *EMBO J* 25:118-128
- Lashkari DA, DeRisi JL, McCusker JH, Namath AF, Gentile C, Hwang SY, Brown PO, & Davis RW (1997) Yeast microarrays for genome wide parallel genetic and gene expression analysis. *Proc Natl Acad Sci U S A* 94:13057-62
- Lee C, Shibata Y, Rao B, Strahl BD, & Lieb JD (2004) Evidence for nucleosome depletion at active regulatory regions genome-wide. *Nat Genet* 36:900-905
- Lee TI et al. (2002) Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* 298:799-804
- Lee YC, Park JM, Min S, Han SJ, & Kim YJ (1999) An activator binding module of yeast RNA polymerase II holoenzyme. *Mol. Cell. Biol* 19:2967-2976
- Lehmann AR (2001) The xeroderma pigmentosum group D (XPD) gene: one gene, two functions, three diseases. *Genes Dev* 15:15-23
- Lescure B, Williamson V, & Sentenac A (1981) Efficient and selective initiation by yeast RNA polymerase B in a dinucleotide-primed reaction. *Nucleic Acids Res* 9:31-45
- Licatalosi DD, Geiger G, Minet M, Schroeder S, Cilli K, McNeil JB, & Bentley DL (2002) Functional interaction of yeast pre-mRNA 3' end processing factors with RNA polymerase II. *Mol. Cell* 9:1101-1111
- Lipshutz RJ, Fodor SP, Gingeras TR, & Lockhart DJ (1999) High density synthetic oligonucleotide arrays. *Nat Genet*
- Liu CL, Kaplan T, Kim M, Buratowski S, Schreiber SL, Friedman N, & Rando OJ (2005) Single-Nucleosome Mapping of Histone Modifications in *S. cerevisiae*. *PLoS Biology* 3:e328 EP -
- Lorenzen K, Vannini A, Cramer P, & Heck AJR (2007) Structural biology of RNA polymerase III: mass spectrometry elucidates subcomplex architecture. *Structure* 15:1237-45
- Lu H, Zawel L, Fisher L, Egly JM, & Reinberg D (1992) Human general transcription factor IIH phosphorylates the C-terminal domain of RNA polymerase II. *Nature* 358:641-645
- Lykke-Andersen S & Jensen TH (2007) Overlapping pathways dictate termination of RNA polymerase II transcription. *Biochimie* 89:1177-1182
- Malagon F, Tong AH, Shafer BK, & Strathern JN (2004) Genetic interactions of DST1 in *Saccharomyces cerevisiae* suggest a role of TFIIIS in the initiation-elongation transition. *Genetics* 166:1215-27
- Margaritis T & Holstege FCP (2008) Poised RNA polymerase II gives pause for thought. *Cell* 133:581-4

- Marmur J & Doty P (1961) Thermal renaturation of deoxyribonucleic acids. *J Mol Biol* 3:585-94
- Martens JA, Laprade L, & Winston F (2004) Intergenic transcription is required to repress the *Saccharomyces cerevisiae* SER3 gene. *Nature* 429:571-4
- Martin DE, Soulard A, & Hall MN (2004) TOR regulates ribosomal protein gene expression via PKA and the Forkhead transcription factor FHL1. *Cell* 119:969-979
- Mason PB & Struhl K (2005) Distinction and relationship between elongation rate and processivity of RNA polymerase II in vivo. *Mol Cell* 17:831-40
- Max T, Sogaard M, & Svejstrup JQ (2007) Hyperphosphorylation of the C-terminal repeat domain of RNA polymerase II facilitates dissociation of its complex with mediator. *J. Biol. Chem* 282:14113-14120
- Maxam AM & Gilbert W (1977) A new method for sequencing DNA. *Proc Natl Acad Sci U S A* 74:560-4
- Medini D, Serruto D, Parkhill J, Relman DA, Donati C, Moxon R, Falkow S, & Rappuoli R (2008) Microbiology in the post-genomic era. *Nat Rev Micro* advanced online publication Available at: <http://dx.doi.org/10.1038/nrmicro1901> [Accédé Mars 20, 2009].
- Meinhart A & Cramer P (2004) Recognition of RNA polymerase II carboxy-terminal domain by 3'-RNA-processing factors. *Nature* 430:223-226
- Meinhart A, Kamenski T, Hoepfner S, Baumli S, & Cramer P (2005) A structural perspective of CTD function. *Genes Dev* 19:1401-1415
- Mitchell AP (1994) Control of meiotic gene expression in *Saccharomyces cerevisiae*. *Microbiol Rev* 58:56-70
- Moqtaderi Z & Struhl K (2004) Genome-wide occupancy profile of the RNA polymerase III machinery in *Saccharomyces cerevisiae* reveals loci with incomplete transcription complexes. *Mol Cell Biol* 24:4118-27
- Morin PE, Awrey DE, Edwards AM, & Arrowsmith CH (1996) Elongation factor TFIIS contains three structural domains: solution structure of domain II. *Proc Natl Acad Sci U S A* 93:10604-8
- Morozova O & Marra MA (2008) Applications of next-generation sequencing technologies in functional genomics. *Genomics* 92:255-64
- Mosley AL, Pattenden SG, Carey M, Venkatesh S, Gilmore JM, Florens L, Workman JL, & Washburn MP (2009) Rtr1 Is a CTD Phosphatase that Regulates RNA Polymerase II during the Transition from Serine 5 to Serine 2 Phosphorylation. *Molecular Cell* 34:168-178
- Mote J, Ghanouni P, & Reines D (1994) A DNA minor groove-binding ligand both potentiates and arrests transcription by RNA polymerase II. Elongation factor SII enables readthrough at arrest sites. *J Mol Biol* 236:725-37

- Muse GW, Gilchrist DA, Nechaev S, Shah R, Parker JS, Grissom SF, Zeitlinger J, & Adelman K (2007) RNA polymerase is poised for activation across the genome. *Nat Genet* 39:1507-1511
- Myers LC, Gustafsson CM, Hayashibara KC, Brown PO, & Kornberg RD (1999) Mediator protein mutations that selectively abolish activated transcription. *Proc. Natl. Acad. Sci. U.S.A* 96:67-72
- Näär AM, Lemon BD, & Tjian R (2001) Transcriptional coactivator complexes. *Annu. Rev. Biochem* 70:475-501
- Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, & Snyder M (2008) The Transcriptional Landscape of the Yeast Genome Defined by RNA Sequencing. *Science* 320:1344-1349
- Nakanishi T, Nakano A, Nomura K, Sekimizu K, & Natori S (1992) Purification, gene cloning, and gene disruption of the transcription elongation factor S-II in *Saccharomyces cerevisiae*. *J Biol Chem* 267:13200-4
- Natori S, Takeuchi K, Takahashi K, & Mizuno D (1973) DNA dependent RNA polymerase from Ehrlich ascites tumor cells. II. Factors stimulating the activity of RNA polymerase II. *J Biochem* 73:879-88
- Neil H, Malabat C, d'Aubenton-Carafa Y, Xu Z, Steinmetz LM, & Jacquier A (2009) Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. *Nature* 457:1038-42
- Nesser NK, Peterson DO, & Hawley DK (2006) RNA polymerase II subunit Rpb9 is important for transcriptional fidelity in vivo. *Proc. Natl. Acad. Sci. U.S.A* 103:3268-3273
- Nonet ML & Young RA (1989) Intragenic and extragenic suppressors of mutations in the heptapeptide repeat domain of *Saccharomyces cerevisiae* RNA polymerase II. *Genetics* 123:715-724
- Oficjalska-Pham D, Harismendy O, Smagowicz WJ, Gonzalez de Peredo A, Boguta M, Sentenac A, & Lefebvre O (2006) General repression of RNA polymerase III transcription is triggered by protein phosphatase type 2A-mediated dephosphorylation of Maf1. *Mol. Cell* 22:623-632
- Olivas WM, Muhlrads D, & Parker R (1997) Analysis of the yeast genome: identification of new non-coding and small ORF-containing RNAs. *Nucleic Acids Res* 25:4619-4625
- Olmsted VK, Awrey DE, Koth C, Shan X, Morin PE, Kazanis S, Edwards AM, & Arrowsmith CH (1998) Yeast transcript elongation factor (TFIIS), structure and function. I: NMR structural analysis of the minimal transcriptionally active region. *J Biol Chem* 273:22589-94

- Opalka N, Chlenov M, Chacon P, Rice WJ, Wriggers W, & Darst SA (2003) Structure and function of the transcription elongation factor GreB bound to bacterial RNA polymerase. *Cell* 114:335-45
- Pallasch CP, Struss A, Munnia A, König J, Steudel W, Fischer U, & Meese E (2005) Autoantibodies against GLEA2 and PHF3 in glioblastoma: tumor-associated autoantibodies correlated with prolonged survival. *Int. J. Cancer* 117:456-459
- Pan G, Aso T, & Greenblatt J (1997) Interaction of elongation factors TFIIS and elongin A with a human RNA polymerase II holoenzyme capable of promoter-specific initiation and responsive to transcriptional activators. *J Biol Chem* 272:24563-71
- van de Peppel J, Kettelarij N, van Bakel H, Kockelkorn TTJP, van Leenen D, & Holstege FCP (2005) Mediator expression profiling epistasis reveals a signal transduction pathway with antagonistic submodules and highly specific downstream targets. *Mol. Cell* 19:511-522
- Phatnani HP & Greenleaf AL (2006) Phosphorylation and functions of the RNA polymerase II CTD. *Genes Dev* 20:2922-2936
- Pluta K, Lefebvre O, Martin NC, Smagowicz WJ, Stanford DR, Ellis SR, Hopper AK, Sentenac A, & Boguta M (2001) Maf1p, a negative effector of RNA polymerase III in *Saccharomyces cerevisiae*. *Mol. Cell. Biol* 21:5031-5040
- Pokholok DK, Hannett NM, & Young RA (2002) Exchange of RNA polymerase II initiation and elongation factors during gene expression in vivo. *Mol Cell* 9:799-809
- Pokholok DK, Harbison CT, Levine S, Cole M, Hannett NM, Lee TI, Bell GW, Walker K, Rolfe PA, Herbolzheimer E, Zeitlinger J, Lewitter F, Gifford DK, & Young RA (2005) Genome-wide Map of Nucleosome Acetylation and Methylation in Yeast. *Cell* 122:517-527
- Powell W & Reines D (1996) Mutations in the second largest subunit of RNA polymerase II cause 6-azauracil sensitivity in yeast and increased transcriptional arrest in vitro. *J. Biol. Chem* 271:6866-6873
- Prakash S & Prakash L (2000) Nucleotide excision repair in yeast. *Mutat. Res* 451:13-24
- Prather DM, Larschan E, & Winston F (2005) Evidence that the elongation factor TFIIS plays a role in transcription initiation at GAL1 in *Saccharomyces cerevisiae*. *Mol Cell Biol* 25:2650-9
- Preker P, Nielsen J, Kammler S, Lykke-Andersen S, Christensen MS, Mapendano CK, Schierup MH, & Jensen TH (2008) RNA Exosome Depletion Reveals Transcription Upstream of Active Human Promoters. *Science* 322:1851-1854
- Qian X, Gozani SN, Yoon H, Jeon CJ, Agarwal K, & Weiss MA (1993) Novel zinc finger motif in the basal transcriptional machinery: three-dimensional NMR studies of the nucleic acid binding domain of transcriptional elongation factor TFIIS. *Biochemistry* 32:9944-59

- Qiu H, Hu C, & Hinnebusch AG (2009) Phosphorylation of the Pol II CTD by KIN28 enhances BUR1/BUR2 recruitment and Ser2 CTD phosphorylation near promoters. *Mol. Cell* 33:752-762
- Radonjic M, Andrau J, Lijnzaad P, Kemmeren P, Kockelkorn TTJP, van Leenen D, van Berkum NL, & Holstege FCP (2005) Genome-wide analyses reveal RNA polymerase II located upstream of genes poised for rapid response upon *S. cerevisiae* stationary phase exit. *Mol Cell* 18:171-83
- Ranish JA, Hahn S, Lu Y, Yi EC, Li X, Eng J, & Aebersold R (2004) Identification of TFB5, a new component of general transcription and DNA repair factor IIH. *Nat. Genet* 36:707-713
- Ranish JA, Yi EC, Leslie DM, Purvine SO, Goodlett DR, Eng J, & Aebersold R (2003) The study of macromolecular complexes by quantitative proteomics. *Nat. Genet* 33:349-355
- Rappaport J, Reinberg D, Zandomeni R, & Weinmann R (1987) Purification and functional characterization of transcription factor SII from calf thymus. Role in RNA polymerase II elongation. *J. Biol. Chem.* 262:5227-5232
- Reese JC (2003) Basal transcription factors. *Curr. Opin. Genet. Dev* 13:114-118
- Reina JH, Azzouz TN, & Hernandez N (2006) Maf1, a new player in the regulation of human RNA polymerase III transcription. *PLoS ONE* 1:e134
- Reinberg D & Roeder R (1987) Factors involved in specific transcription by mammalian RNA polymerase II. Transcription factor IIS stimulates elongation of RNA chains. *J. Biol. Chem.* 262:3331-3337
- Reines D (1992) Elongation factor-dependent transcript shortening by template-engaged RNA polymerase II. *J. Biol. Chem.* 267:3795-3800
- Reines D, Chamberlin M, & Kane C (1989) Transcription elongation factor SII (TFIIS) enables RNA polymerase II to elongate through a block to transcription in a human gene in vitro. *J. Biol. Chem.* 264:10799-10809
- Reines D, Ghanouni P, Gu W, Mote J, & Powell W (1993) Transcription elongation by RNA polymerase II: mechanism of SII activation. *Cell Mol Biol Res* 39:331-8
- Reines D & Mote J (1993) Elongation factor SII-dependent transcription by RNA polymerase II through a sequence-specific DNA-binding protein. *Proceedings of the National Academy of Sciences of the United States of America* 90:1917-1921
- Ren B, Robert F, Wyrick JJ, Aparicio O, Jennings EG, Simon I, Zeitlinger J, Schreiber J, Hannett N, Kanin E, Volkert TL, Wilson CJ, Bell SP, & Young RA (2000) Genome-wide location and function of DNA binding proteins. *Science* 290:2306-9
- Roberts DN, Stewart AJ, Huff JT, & Cairns BR (2003) The RNA polymerase III transcriptome revealed by genome-wide localization and activity-occupancy relationships. *Proc Natl Acad Sci U S A* 100:14695-700

- Roberts S, Miller SJ, Lane WS, Lee S, & Hahn S (1996) Cloning and functional characterization of the gene encoding the TFIIB90 subunit of RNA polymerase III transcription factor TFIIB. *J. Biol. Chem* 271:14903-14909
- Robertson G, Hirst M, Bainbridge M, Bilenky M, Zhao Y, Zeng T, Euskirchen G, Bernier B, Varhol R, Delaney A, Thiessen N, Griffith OL, He A, Marra M, Snyder M, & Jones S (2007) Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat Methods* 4:651-7
- Rondón AG, Jimeno S, García-Rubio M, & Aguilera A (2003) Molecular evidence that the eukaryotic THO/TREX complex is required for efficient transcription elongation. *J. Biol. Chem* 278:39037-39043
- Rosonina E, Willis IM, & Manley JL (2009) Sub1 functions in osmoregulation and in transcription by both RNA polymerases II and III. *Mol. Cell. Biol* 29:2308-2321
- Ross-Macdonald P, Coelho PS, Roemer T, Agarwal S, Kumar A, Jansen R, Cheung KH, Sheehan A, Symoniatis D, Umansky L, Heidtman M, Nelson FK, Iwasaki H, Hager K, Gerstein M, Miller P, Roeder GS, & Snyder M (1999) Large-scale analysis of the yeast genome by transposon tagging and gene disruption. *Nature* 402:413-8
- Roy R, Schaeffer L, Humbert S, Vermeulen W, Weeda G, & Egly JM (1994) The DNA-dependent ATPase activity associated with the class II basic transcription factor BTF2/TFIIB. *J. Biol. Chem* 269:9826-9832
- Rudd MD, Izban MG, & Luse DS (1994) The active site of RNA polymerase II participates in transcript cleavage within arrested ternary complexes. *Proc Natl Acad Sci U S A* 91:8057-61
- Rüth J, Conesa C, Dieci G, Lefebvre O, Düsterhöft A, Ottonello S, & Sentenac A (1996) A suppressor of mutations in the class III transcription system encodes a component of yeast TFIIB. *EMBO J* 15:1941-1949
- Ryu S, Zhou S, Ladurner AG, & Tjian R (1999) The transcriptional cofactor complex CRSP is required for activity of the enhancer-binding protein Sp1. *Nature* 397:446-450
- Sanger F & Coulson AR (1975) A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J Mol Biol* 94:441-8
- Sanger F, Nicklen S, & Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* 74:5463-7
- Sastry SS & Ross BM (1997) Nuclease activity of T7 RNA polymerase and the heterogeneity of transcription elongation complexes. *J. Biol. Chem* 272:8644-8652
- Saunders A, Core LJ, & Lis JT (2006) Breaking barriers to transcription elongation. *Nat Rev Mol Cell Biol* 7:557-567

- Sawadogo M, Huet J, & Fromageot P (1980) Similar binding site for P37 factor on yeast RNA polymerases A and B. *Biochem. Biophys. Res. Commun* 96:258-264
- Sawadogo M, Lescure B, Sentenac A, & Fromageot P (1981) Native deoxyribonucleic acid transcription by yeast RNA polymerase--P37 complex. *Biochemistry* 20:3542-3547
- Sawadogo M, Sentenac A, & Fromageot P (1980) Interaction of a new polypeptide with yeast RNA polymerase B. *J Biol Chem* 255:12-5
- Schaeffer L, Roy R, Humbert S, Moncollin V, Vermeulen W, Hoeijmakers JH, Chambon P, & Egly JM (1993) DNA repair helicase: a component of BTF2 (TFIIH) basic transcription factor. *Science* 260:58-63
- Schawalder SB, Kabani M, Howald I, Choudhury U, Werner M, & Shore D (2004) Growth-regulated recruitment of the essential yeast ribosomal protein gene activator Ifh1. *Nature* 432:1058-1061
- Schena M, Shalon D, Davis RW, & Brown PO (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270:467-70
- Schnapp G, Graveley BR, & Grummt I (1996) TFIIS binds to mouse RNA polymerase I and stimulates transcript elongation and hydrolytic cleavage of nascent rRNA. *Mol. Gen. Genet* 252:412-419
- Schneider DA, French SL, Osheim YN, Bailey AO, Vu L, Dodd J, Yates JR, Beyer AL, & Nomura M (2006) RNA polymerase II elongation factors Spt4p and Spt5p play roles in transcription elongation by RNA polymerase I and rRNA processing. *Proc. Natl. Acad. Sci. U.S.A* 103:12707-12712
- Schones DE & Zhao K (2008) Genome-wide approaches to studying chromatin modifications. *Nat Rev Genet* 9:179-91
- Schramm L & Hernandez N (2002) Recruitment of RNA polymerase III to its target promoters. *Genes Dev* 16:2593-2620
- Schroeder SC, Schwer B, Shuman S, & Bentley D (2000) Dynamic association of capping enzymes with transcribing RNA polymerase II. *Genes Dev* 14:2435-2440
- Schultz P, Fribourg S, Poterszman A, Mallouh V, Moras D, & Egly JM (2000) Molecular structure of human TFIIH. *Cell* 102:599-607
- Seila AC, Calabrese JM, Levine SS, Yeo GW, Rahl PB, Flynn RA, Young RA, & Sharp PA (2008) Divergent Transcription from Active Promoters. *Science* 322:1849-1851
- Sekimizu K, Kobayashi N, Mizuno D, & Natori S (1976) Purification of a factor from Ehrlich ascites tumor cells specifically stimulating RNA polymerase II. *Biochemistry* 15:5064-70



- Serizawa H, Conaway RC, & Conaway JW (1992) A carboxyl-terminal-domain kinase associated with RNA polymerase II transcription factor delta from rat liver. *Proc. Natl. Acad. Sci. U.S.A* 89:7476-7480
- Shaaban SA, Bobkova EV, Chudzik DM, & Hall BD (1996) In vitro analysis of elongation and termination by mutant RNA polymerases with altered termination behavior. *Mol. Cell. Biol* 16:6468-6476
- Shilatifard A, Lane WS, Jackson KW, Conaway RC, & Conaway JW (1996) An RNA polymerase II elongation factor encoded by the human ELL gene. *Science* 271:1873-1876
- Shilatifard A, Conaway RC, & Conaway JW (2003) The RNA polymerase II elongation complex. *Annu Rev Biochem* 72:693-715
- Sluder AE, Greenleaf AL, & Price DH (1989) Properties of a *Drosophila* RNA polymerase II elongation factor. *J Biol Chem* 264:8963-9
- Solomon MJ & Varshavsky A (1985) Formaldehyde-mediated DNA-protein crosslinking: a probe for in vivo chromatin structures. *Proc Natl Acad Sci U S A* 82:6470-4
- Southern EM (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol* 98:503-17
- Spellman PT, Sherlock G, Zhang MQ, Iyer VR, Anders K, Eisen MB, Brown PO, Botstein D, & Futcher B (1998) Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol Biol Cell* 9:3273-97
- Stebbins CE, Borukhov S, Orlova M, Polyakov A, Goldfarb A, & Darst SA (1995) Crystal structure of the GreA transcript cleavage factor from *Escherichia coli*. *Nature* 373:636-640
- Stein A & Aloy P (2008) A molecular interpretation of genetic interactions in yeast. *FEBS Lett* 582:1245-1250
- Steinmetz EJ, Warren CL, Kuehner JN, Panbehi B, Ansari AZ, & Brow DA (2006) Genome-wide distribution of yeast RNA polymerase II and its control by Sen1 helicase. *Mol Cell* 24:735-46
- Struss AK, Romeike BF, Munnia A, Nastainczyk W, Steudel WI, König J, Ohgaki H, Feiden W, Fischer U, & Meese E (2001) PHF3-specific antibody responses in over 60% of patients with glioblastoma multiforme. *Oncogene* 20:4107-4114
- Sun ZW, Tessmer A, & Hampsey M (1996) Functional interaction between TFIIB and the Rpb9 (Ssu73) subunit of RNA polymerase II in *Saccharomyces cerevisiae*. *Nucleic Acids Res* 24:2560-2566
- Taira Y, Kubo T, & Natori S (1998) Molecular cloning of cDNA and tissue-specific expression of the gene for SII-K1, a novel transcription elongation factor SII. *Genes Cells* 3:289-296

- Takagi Y, Calero G, Komori H, Brown JA, Ehrensberger AH, Hudmon A, Asturias F, & Kornberg RD (2006) Head module control of mediator interactions. *Mol. Cell* 23:355-364
- Thiebaut M, Colin J, Neil H, Jacquier A, Séraphin B, Lacroute F, & Libri D (2008) Futile cycle of transcription initiation and termination modulates the response to nucleotide shortage in *S. cerevisiae*. *Mol. Cell* 31:671-682
- Thomas MJ, Platas AA, & Hawley DK (1998) Transcriptional fidelity and proofreading by RNA polymerase II. *Cell* 93:627-637
- Thomas MC & Chiang C (2006) The general transcription machinery and general cofactors. *Crit. Rev. Biochem. Mol. Biol* 41:105-178
- Thuillier V, Brun I, Sentenac A, & Werner M (1996) Mutations in the alpha-amanitin conserved domain of the largest subunit of yeast RNA polymerase III affect pausing, RNA cleavage and transcriptional transitions. *EMBO J* 15:618-629
- Triezenberg SJ (1995) Structure and function of transcriptional activation domains. *Curr. Opin. Genet. Dev* 5:190-196
- Tschochner H (1996) A novel RNA polymerase I-dependent RNase activity that shortens nascent transcripts from the 3' end. *Proc. Natl. Acad. Sci. U.S.A* 93:12914-12919
- Ubukata T, Shimizu T, Adachi N, Sekimizu K, & Nakanishi T (2003) Cleavage, but not read-through, stimulation activity is responsible for three biologic functions of transcription elongation factor S-II. *J Biol Chem* 278:8580-5
- Uetz P et al. (2000) A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* 403:623-7
- Umehara T, Kida S, Yamamoto T, & Horikoshi M (1995) Isolation and characterization of a cDNA encoding a new type of human transcription elongation factor S-II. *Gene* 167:297-302
- Upadhyaya R, Lee J, & Willis IM (2002) MafI is an essential mediator of diverse signals that repress RNA polymerase III transcription. *Mol. Cell* 10:1489-1494
- Venters BJ & Pugh BF (2009) A canonical promoter organization of the transcription machinery and its regulators in the *Saccharomyces* genome. *Genome Res* 19:360-71
- Wade JT, Hall DB, & Struhl K (2004) The transcription factor Ifh1 is a key regulator of yeast ribosomal protein genes. *Nature* 432:1054-1058
- Wade JT & Struhl K (2008) The transition from transcriptional initiation to elongation. *Curr Opin Genet Dev* 18:130-6
- Walhout AJ, Boulton SJ, & Vidal M (2000) Yeast two-hybrid systems and protein interaction mapping projects for yeast and worm. *Yeast* 17:88-94
- Wang D & Hawley DK (1993) Identification of a 3'-->5' exonuclease activity associated with human RNA polymerase II. *Proc Natl Acad Sci U S A* 90:843-7

- Wang D, Bushnell DA, Huang X, Westover KD, Levitt M, & Kornberg RD (2009) Structural basis of transcription: backtracked RNA polymerase II at 3.4 angstrom resolution. *Science* 324:1203-1206
- Weaver ZA & Kane CM (1997) Genomic characterization of a testis-specific TFIIIS (TCEA2) gene. *Genomics* 46:516-9
- Werner M, Chaussivert N, Willis IM, & Sentenac A (1993) Interaction between a complex of RNA polymerase III subunits and the 70-kDa component of transcription factor IIIB. *J. Biol. Chem* 268:20721-20724
- Werner M, Hermann-Le Denmat S, Treich I, Sentenac A, & Thuriaux P (1992) Effect of mutations in a zinc-binding domain of yeast RNA polymerase C (III) on enzyme function and subunit association. *Mol. Cell. Biol* 12:1087-1095
- Wery M, Shematorova E, Van Driessche B, Vandenhoute J, Thuriaux P, & Van Mullem V (2004) Members of the SAGA and Mediator complexes are partners of the transcription elongation factor TFIIIS. *EMBO J* 23:4232-42
- Westover KD, Bushnell DA, & Kornberg RD (2004) Structural basis of transcription: nucleotide selection by rotation in the RNA polymerase II active center. *Cell* 119:481 - 9
- White RJ (2004) RNA polymerase III transcription and cancer. *Oncogene* 23:3208-3216
- White RJ (2008) RNA polymerases I and III, non-coding RNAs and cancer. *Trends Genet* 24:622-629
- Whitehall SK, Bardeleben C, & Kassavetis GA (1994) Hydrolytic cleavage of nascent RNA in RNA polymerase III ternary transcription complexes. *J. Biol. Chem* 269:2299-2306
- Wickner RB, Edskes HK, Shewmaker F, & Nakayashiki T (2007) Prions of fungi: inherited structures and biological roles. *Nat Rev Microbiol* 5:611-8
- Wiest DK, Wang D, & Hawley DK (1992) Mechanistic studies of transcription arrest at the adenovirus major late attenuation site. Comparison of purified RNA polymerase II and washed elongation complexes. *J Biol Chem* 267:7733-44
- Wind M & Reines D (2000) Transcription elongation factor SII. *Bioessays* 22:327-36
- Winzeler EA et al. (1999) Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* 285:901-6
- Wolfe KH & Shields DC (1997) Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* 387:708-713
- Wood A & Shilatifard A (2006) Bur1/Bur2 and the Ctk complex in yeast: the split personality of mammalian P-TEFb. *Cell Cycle* 5:1066-1068

- Wu J, Awrey DE, Edwards AM, Archambault J, & Friesen JD (1996) In vitro characterization of mutant yeast RNA polymerase II with reduced binding for elongation factor TFIIS. *Proc Natl Acad Sci U S A* 93:11552-7
- Wu JQ & Snyder M (2008) RNA polymerase II stalling: loading at the start prepares genes for a sprint. *Genome Biol* 9:220
- Wu X, Rossetini A, & Hanes SD (2003) The ESS1 prolyl isomerase and its suppressor BYE1 interact with RNA pol II to inhibit transcription elongation in *Saccharomyces cerevisiae*. *Genetics* 165:1687-1702
- Wyers F, Rougemaille M, Badis G, Rousselle J, Dufour M, Boulay J, Régnault B, Devaux F, Namane A, Séraphin B, Libri D, & Jacquier A (2005) Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. *Cell* 121:725-37
- Xiao Z, McGrew JT, Schroeder AJ, & Fitzgerald-Hayes M (1993) CSE1 and CSE2, two new genes required for accurate mitotic chromosome segregation in *Saccharomyces cerevisiae*. *Mol. Cell. Biol* 13:4691-4702
- Xiao ZX & Fitzgerald-Hayes M (1995) Functional interaction between the CSE2 gene product and centromeres in *Saccharomyces cerevisiae*. *J. Mol. Biol* 248:255-263
- Xu Q, Nakanishi T, Sekimizu K, & Natori S (1994) Cloning and identification of testis-specific transcription elongation factor S-II. *J. Biol. Chem* 269:3100-3103
- Xu Z, Wei W, Gagneur J, Perocchi F, Clauder-Munster S, Camblong J, Guffanti E, Stutz F, Huber W, & Steinmetz LM (2009) Bidirectional promoters generate pervasive transcription in yeast. *Nature* 457:1033-1037
- Yuan G, Liu Y, Dion MF, Slack MD, Wu LF, Altschuler SJ, & Rando OJ (2005) Genome-Scale Identification of Nucleosome Positions in *S. cerevisiae*. *Science* 309:626-630
- Yudkovsky N, Ranish JA, & Hahn S (2000) A transcription reinitiation intermediate that is stabilized by activator. *Nature* 408:225-229
- Zeitlinger J, Stark A, Kellis M, Hong J, Nechaev S, Adelman K, Levine M, & Young RA (2007) RNA polymerase stalling at developmental control genes in the *Drosophila melanogaster* embryo. *Nat Genet* 39:1512-1516
- Zhu X, Wirén M, Sinha I, Rasmussen NN, Linder T, Holmberg S, Ekwall K, & Gustafsson CM (2006) Genome-wide occupancy profile of mediator and the Srb8-11 module reveals interactions with coding regions. *Mol Cell* 22:169-78



## **ANNEXES**



## Protocole de Recombineering

But : Insérer par recombineering une étiquette en N-terminal d'un gène dans un locus de souris.

### I) Constructions *in silico*

Localiser le gène d'intérêt avec ENSEMBL.

Dans Detailed view, sélectionner dans DAS source l'option [129S7/AB2.2 clones](#) pour commander le BAC. Comme l'étiquette doit être insérée en N-ter choisir 2 BAC qui pourront contenir 10 kb de part et d'autre du N-ter. Les commander (Geneservice).

Extraire sous forme FASTA la séquence -10 kb - « N-ter (codon ATG) »- +10 kb et la mettre dans DNA strider ou ApE.

Repérer précisément le codon ATG et adopter une stratégie pour le génotypage des cellules ES. Ce génotypage se fera par Southern blot. Il faut choisir une enzyme de restriction qui coupe idéalement 2 fois (une à l'extérieur des bras distaux, et l'autre à l'intérieur). Cette enzyme doit être une de celles utilisées par le laboratoire pour les Southern blot (ne doit pas être inhibée par CpG méthylation).

BamHI, NcoI, EcoRI, XhoI, XbaI, HindIII et SpeI sont apportés par la construction. Placer la sonde à l'extérieur des bras distaux 5' ou 3'. Cette sonde ne doit pas être disposée dans des régions répétées. Pour le vérifier utiliser le programme Repeat Masker et sélectionner le génome de la souris.

Choisir les amorces PCR adéquates avec Primer3 Input pour générer une sonde de 600-700 pb.

**IMPORTANT : Si possible choisir 2 stratégies: une avec une sonde en 5' et l'autre avec une sonde en 3' pour pouvoir voir un shift de 2 kb quand la cassette Néo sera excisée.**

**Tenir compte du placement des bras distaux (la sonde doit être à l'extérieur).**

Le profil entre le WT et le mutant en Southern blot devra être assez différent :

Ex : Si pour le WT on obtient une bande à 10 kb, dans l'idéal on doit avoir une bande à 5 kb pour le mutant.

Une fois la sonde placée, disposer les bras distaux 5' et 3'. Ils doivent être idéalement placés à respectivement -5 kb et +5 kb du Stop et comprendre entre 250-500 pb. Dans tous les cas s'assurer qu'il y a 10 kb entre les bras distaux. On peut avoir un bras court à 3kb et un bras long à 7kb.

Choisir les amorces adéquates avec Primer3 Input.

**Comme les bras distaux seront amplifiés par PCR et clonés dans le "retrieval plasmid" (en NotI / SpeI pour le bras distal 5' et SpeI / BamHI pour le bras distal 3') vérifier qu'il n'y a pas de sites de restriction NotI / SpeI dans le 5' et Spe I / Bam HI dans le 3'.**

Ajouter dans l'amorce FW du bras distal 5' le site NotI, et dans l'amorce Rev le site SpeI.

Ajouter dans l'amorce FW du bras distal 3' le site SpeI, et dans l'amorce Rev le site BamHI.

Note: lors de choix des amorces, toujours choisir des séquences de restrictions les plus longues possibles (optimisation des digestions à proximité de la fin d'un fragment d'ADN).

Les bras d'homologie proximaux doivent être compris entre 250-500 pb. L'amorce Rev du bras proximal 5' comprendra les 20 nucléotides avant le codon ATG. L'amorce FW du bras proximal 3' comprendra les 20 nucléotides suivants codon ATG compris.



Ajouter dans l'amorce FW du bras proximal 5' le site Sall, et dans l'amorce Rev le site EcoRI.

Ajouter dans l'amorce FW du bras proximal 3' le site HindIII, et dans l'amorce Rev le site NotI.

**IMPORTANT : Le bras proximal 5' ne devra pas contenir de sites Sall et EcoRI. Le 3' ne devra pas posséder de sites HindIII et NotI.**

Rapatrier dans DNA strider ou ApE les séquences des bras d'homologie proximaux et distaux, des sondes pour toutes les constructions, et les séquences de 20 kb flanquant l'ATG du gène d'intérêt.

Construire la carte du "Gap repaired plasmid", du "retrieval plasmid", du "minitargeting vector", du "retrieval plasmid" et du "neotargeted plasmid".

## II) Insertion du BAC dans la souche SW102 et Vérifications

Préculture des bactéries renfermant les BAC dans 5 mL LB/Chloramphénicol 12,5µg/ml final. Mettre les BAC en collection.

Purifier les 2 BAC sélectionnés (Miniprep d'ADN plasmidique en utilisant les solutions P1, P2, N3 du kit QIAGEN, puis précipitation à l'EtOH au lieu d'utiliser les colonnes) et resuspendre dans 50µl d'eau stérile.

Doser l'ADN pour vérifier sa pureté et la quantité. (Faire une dilution au 1/100 et doser à 260 nm).

Vérifier sur gel le profil des BAC préalablement digérés par une, deux voir 3 enzymes de restriction (une, deux ou 3 digestions indépendantes). PS : j'ai vérifié mes BAC avec HindIII, EcoRI et/ou KpnI.

Les digestions doivent être réalisées avec une quantité suffisante d'ADN (au moins 20 à 30 µl) pendant 3hrs (ou overnight) à 37°C.

Migration sur gel agarose 1% ON à 40 V. (de 18h30 à 9h00). Le lendemain il est parfois utile d'incuber le gel dans un bain de TBE1X + BET avant de prendre une photo.

Transformation par électroporation des cellules **SW102**

1. Pré-culture de 5 ml LB ON à 30°C.
2. Mettre de l'eau, des falcons 50, des eppendorfs et des cuvettes à 4°C.
3. Diluer au 50<sup>ème</sup> (500µL dans 25 mL) et incuber 3-5h jusqu'à atteindre une DO<sub>600</sub> = 0,6. 25 mL de culture correspond à 2 tubes de cellules compétentes.
4. Incuber les cellules dans la glace pendant 2 min puis les transférer (12,5 ml) dans 2 falcon50 préalablement à 4°C.
5. Centrifuger à 4°C, 5000 rpm pdt 5 min.
6. Enlever le surnageant et égoutter sur papier, puis ajouter 5 ml d'eau à 4°C en gardant le tube dans la glace. Resuspendre le culot en agitant doucement (en faisant un cercle) le falcon placé dans la glace pdt 5 min. Quand les cellules sont resuspendues, compléter à 15 mL avec de l'eau glacée puis inverser quelques fois.
7. Centrifuger à 4°C, 5000 rpm pdt 5 min.
8. Refaire les étapes 6 et 7 (la resuspension est plus facile cette fois). Enlever le surnageant en inversant le tube sur du papier (Ne pas perdre le culot). Resuspendre le

culot dans le volume résiduel du tube et stocker les cellules compétentes à 4°C (env 200µL par falcon).

9. Transférer 50µL de cellules électrocompétentes dans un eppendorf préalablement mis à 4°C et mélanger avec 2 µL d'ADN. Pour un BAC utiliser 1-5µg d'ADN (ici environ 10 µg). Transformer par électroporation (1750 V). Vérifier le "time constant". Il doit être aux alentours de 5-6mv.
10. Transférer les cellules dans 1 mL de LB puis incubé à 30°C pdt 1h.
11. Etaler les bactéries (30 µL 150µL et le reste) sur le milieu de sélection ici : LB/Chloramphénicol. Centrifuger à 5000 rpm et enlever le surnageant pour le reste. Resuspendre dans 150µl et étaler.
12. Incuber à 30°C pour 16-24h. NB : Inclure un contrôle sans BAC et l'étaler sur une boîte LB et une LB/Chloramphénicol.

Le lendemain :

Les colonies sont très petites. Il est préférable de les laisser une bonne partie de la journée.

Repiquer 5 clones dans 3 ml de LB/Cm ON à 30°C.

NB: Si les transformations ne marchent pas bien, purifier les BAC au phénol-chloroforme avant la transformation. De plus, il est préférable de travailler avec des préparations de BAC fraîches.

Repiquer 2 à 4 clones par BAC, les mettre en collection et purifier les BAC à partir de la souche SW102.

Vérifier sur gel le profil des BAC préalablement digérés par une, deux voire 3 enzymes de restriction. Digérer le BAC original (issu de la souche *E. coli* initiale) et 2 BAC issus de la souche SW102.

Comparer les profils et valider le BAC.

Vérifier les BAC par PCR.

### III) Vérification des couples d'amorces pour PCR (Sonde, et bras d'homologie distaux et proximaux).

Mix réactionnel :

1µl ADN BAC

1µl Amorces FW 10µM

1µl Amorces Rev 10µM

2,5µl Tampon 10X

0,2µl Taq

1µl MgCl<sub>2</sub> 50 mM

1µl dNTPs 10 mM

17,3 µl eau

Programme PCR :

PCR Bac 60 avec

2' à 94°C

15'' à 94°C

30'' à 60°C

45'' à 72°C

4' à 72°C

} 30 cycles

Faire migrer 5-10 µl de la PCR dans un gel d'agarose 1%. Stocker à 4°C le reste de la PCR (important pour la suite).

**Amplification par PCR des bras d'homologie distaux et proximaux et sonde à partir du BAC issu de la souche initiale d'*E. coli* et du BAC transformé dans souche SW102 (vérification du BAC).**

**Refaire les PCR mais cette fois faire 4 tubes de 100 µl pour chaque produit de PCR (pour générer suffisamment de matériel pour le clonage). Faire un contrôle négatif pour chaque PCR (chaque couple d'amorces).**

Faire migrer 4µl PCR avec à côté 5-10µl de la PCR ayant servi à contrôler les amorces (pour vérifier la taille de l'amplicon et ainsi voir s'il n'y a pas eu d'inversions).

Purifier le reste des produits PCR sur colonne PCR Purification Kit (QIAGEN). Reprendre les échantillons dans 30µl d'eau<sub>dd</sub>.

NB: Si les PCR ne marchent pas bien, utiliser l'enzyme KOD Hot Start Polymerase (Novagen), ajouter du DMSO 5% et essayer de faire une PCR "touchdown".

#### IV) Construction du Retrieval Plasmid: Clonage des bras distaux 5' et 3' dans pL253

##### 1. Digestion du bras distal 5' par NotI / SpeI

Mix réactionnel :

30µL PCR  
1µl NotI  
1µl SpeI  
3,5µl NEB2

##### 2. Digestion du bras distal 3' par SpeI / BamHI

Mix réactionnel :

30µL PCR  
1µl SpeI  
1µl BamHI  
3,5µl React 4

##### 3. Digestion de pL253

Digérer pL253 (au moins 20µg) par NotI / BamHI. Tampon commun : React 3

Si possible, inactiver les enzymes utilisées pour les digestions.

En contrôle pour vérifier que les enzymes de restriction sont fonctionnelles digérer séparément 2µg de pL253 avec SpeI, BamHI et NotI.

Toutes les digestions sont incubées à 37°C pendant 2h30 (ou overnight)

##### 4. Purification sur gel des digestions (Kit Gel Extraction QIAGEN)

##### 5. Traitement du plasmide digéré à la phosphatase alcaline

16 µl plasmide digéré

2 µl tampon NEB3

1 µl CIP

Incuber 30 min à 37°C.

Ajouter 1 µl CIP.

Incuber 30 min à 37°C.

Inactiver l'enzyme 30 min à 65°C.

Extraction phénol chloroforme. Précipitation à l'EtOH, culot repris dans 8 µl eau.

6. Quantification des produits digérés et purifiés sur gel au nanodrop  
**Formule :  $\text{ng}/\mu\text{l} * 10^3 / (635 * \text{nb bases (de l'échantillon)}) = \text{pmoles}/\mu\text{l}$ .**

7. Ligation des miniarms dans pL253

Mix réactionnel :

Tampon 5X 4µl

pL253 (NotI / BamHI) 0,1 pmole

bras distal 5' (NotI / SpeI) 0,5 pmole

bras distal 3' (SpeI / BamHI) 0,5 pmole

Taq DNA Ligase 1µl

Qsp 20µl Eau

Incuber ON à 16°C

Faire un contrôle sans insert.

8. Transformation des produits de ligation  
Transformation de préférence dans cellules DH10B electrocompetentes.  
Étaler sur des boîtes LB agar / **Amp** 100µg/ml final.

Inclure un contrôle sans ADN et étaler 100µL sur LB et 100µl sur LB/**Amp**.

9. Purification de l'ADN des clones qui ont poussé  
Piquer 8 clones par construction et les ensemercer dans 3 ml de LB/Amp. Incuber ON à 37°C.  
Purification ADN plasmidique

10. Vérification du clonage du retrieval plasmid  
Digestion pour sortir l'insert  
4µl ADN  
0,3µl Enz I  
0,3µl Enz II  
2µl Tampon  
13,4µl Eau  
Digérer 2-3hrs à 37°C dans un bain marie.  
Migrer sur un gel agarose 1%.

Mettre en collection 2 clones positifs pour chaque construction.

NB : on peut également digérer pL253 et le retrieval plasmid par SpeI et les faire migrer sur gel.

### 11. Séquençage des produits de ligation

On séquence deux clones de chaque construction sur le brin sens et le brin anti-sens.

## III) Construction du Gap repaired Plasmid (GRP)

### A) Digestion du retrieval plasmid

Mix réactionnel :

2µg ADN miniprep (dans l'idéal, mais marche avec moins)

3µl Tampon React 4

1µl SpeI

Eau Qsp 30µl

Digestion 3h30 (ou overnight) à 37°C.

Vérifier sur gel la digestion en déposant 3µl sur un gel d'agarose 1%.

Purifier sur colonne QIAGEN les plasmides digérés par SpeI (il faut que la digestion du plasmide soit totale). Si la digestion n'est pas totale et que l'on a suffisamment de matériel, une purification sur gel peut être envisagée.

### B) Transformation dans SW102 renfermant les BAC des retrieval plasmid correspondants: induction de la recombinaison homologue.

1. Pré-culture de 5 ml LB/Cm ON à 30-32 °C.
2. Mettre de l'eau, des falcons 50, des eppendorfs et des cuvettes à 4°C.
3. Diluer au 50<sup>ème</sup> (500µL dans 25 mL) et incuber 3-5h jusqu'à atteindre une DO<sub>600</sub> =0,6.
4. Incuber les cellules dans la glace pdt 2 min puis les transférer (12,5 mL) dans 2 falcons 50 préalablement refroidis.
5. Traiter pendant 15' un échantillon à 30°C et un autre à 42°C pour induire la production des protéines de recombinaison. Agiter manuellement !!!!
6. Refroidir les cellules dans de l'eau glacée pendant 5 min en agitant.
7. Centrifuger à 4°C, 5000 rpm pdt 5 min.
8. Enlever le surnageant et égoutter sur papier, puis ajouter 5 ml d'eau à 4°C en gardant le tube dans la glace. Resuspendre le culot en agitant doucement (en faisant un cercle) le falcon placé dans la glace pdt 5 min. Quand les cellules sont resuspendues, compléter à 15 mL avec de l'eau glacée puis inverser quelques fois.
9. Centrifuger à 4°C, 5000 rpm pdt 5 min.
10. Refaire les étapes 6 et 7 (la resuspension est plus facile cette fois).
11. Enlever le surnageant en inversant le tube sur du papier (Ne pas perdre le culot). Resuspendre le culot dans le volume résiduel du tube et stocker les cellules compétentes à 4°C (env 100µL par falcon).
12. Transférer les cellules électrocompétentes dans un eppendorf préalablement mis à 4°C et mélanger idéalement 50 ng de Retrieval plasmid digéré par SpeI.
13. Transformer par électroporation à 1750 V. Transférer les cellules dans 1 mL de LB puis incuber à 30-32°C pdt 1h.

14. Etaler les bactéries (1  $\mu$ L 100 $\mu$ L et le reste) sur le milieu de sélection (LB/Amp pour sélectionner le retrieval plasmid).

Incuber à 30°C pour 16-24h.

Le lendemain :

Préculture de 6 clones dans 3mL de LB/Amp ON à 30°C.

Préparation des plasmides par Miniprep (QIAGEN)

### **C) Vérification du Gap repaired Plasmid : GRP (vérification de l'insertion des 10 kb)**

Définir une stratégie de digestion entre le GRP et le Retrieval Plasmid.

Essayer d'avoir au max 3 bandes. Utiliser soit une simple ou une double digestion.

Idéalement on linéarise le RP et on obtient 2 bandes pour le GRP.

Digérer 3 $\mu$ l d'ADN dans 20 $\mu$ l final.

Mettre en collection le Gap repaired plasmid validé.

## **IV) Construction du mini-targeting vector (MTV)**

Clonage des bras d'homologie proximaux 5' et 3' dans pL452+tag

**IMPORTANT** : En fonction des sites de restriction présents dans les bras proximaux on doit choisir par quel bras d'homologie on doit cloner dans pL452+tag.

### **A) Clonage du bras proximal 5' dans pL452-tag**

1. Digestion du bras proximal 5' par Sall / EcoRI
2. Digestion de pL452+tag par Sall /EcoRI (idéalement partir d'environ 20  $\mu$ g de plasmide)
3. Purifier sur gel les produits de digestion.
4. Ligation et clonage idem précédemment (déphosphorylation du vecteur digéré)
5. Vérification de l'insertion du bras d'homologie 5'.
6. Mettre en collection le plasmide pL452+tag+bras proximal 5' vérifié.

### **B) Clonage du bras proximal 3' dans pL452+tag+bras proximal 5'**

1. Digestion du bras proximal 3' par NotI / HindIII
2. Digestion du pL452+tag+bras proximal 5' par NotI / HindIII
3. Suivre le protocole habituel de clonage.

4. Vérification de l'insertion du bras d'homologie 5'
5. Vérification de la construction de la construction finale par digestion et séquençage.
6. Mettre en collection le mini-targeting vector (MTV) vérifié.

**IMPORTANT : vérifier par séquençage que le TAG est en phase et que les sites de restriction pour le géotypage sont bien présents.**

## V) Construction du neo-targeted plasmid (NTP)

### A) Préparation du MTV

#### B) Extraction de la mini-targeting cassette

1. Digestion MTV par NotI / Sall

Mix réactionnel :

20µg MTV

2µl NotI

2µl Sall

3µl React 3

Digestion 2h30-3h ou ON à 37°C

2. Purification sur gel

Il n'est souvent pas possible de purifier les 2 bandes issues de la digestion. Cette purif sur gel est faite pour éliminer le MTV qui n'a pas été digéré.

3. Quantification des produits.

#### C) Transformation de la mini-targeting cassette dans la souche SW102 contenant le **Gap Repaired Plasmid**.

1. Pré-culture de 5 ml LB ON à 30-32 °C.
2. Mettre de l'eau, des falcons 50, des eppendorfs et des cuvettes à 4°C.
3. Diluer au 50<sup>ème</sup> (500µL dans 25 mL) et incubé 3-5h jusqu'à atteindre une DO<sub>600</sub> = 0,6.
4. Incuber les cellules dans la glace pdt 2 min puis les transférer (12,5 mL) dans 2 falcons 50 préalablement mis à 4°C.
5. Traiter pendant 15' un échantillon à 30°C et un autre à 42°C pour induire la production des protéines de recombinaison. Agiter manuellement !!!!
6. Refroidir les cellules dans de l'eau glacée pendant 5 min.
7. Centrifuger à 4°C, 5000 rpm pdt 5 min.
8. Enlever le surnageant et égoutter sur papier, puis ajouter 5 ml d'eau à 4°C en gardant le tube dans la glace. Resuspendre le culot en agitant doucement (en faisant un cercle) le falcon placé dans la glace pdt 5 min. Quand les cellules sont resuspendues, compléter à 15 mL avec de l'eau glacée puis inverser quelques fois.
9. Centrifuger à 4°C, 5000 rpm pdt 5 min.
10. Refaire les étapes 6 et 7 (la resuspension est plus facile cette fois).

11. Enlever le surnageant en inversant le tube sur du papier (Ne pas perdre le culot). Resuspendre le culot dans le volume résiduel du tube et stocker les cellules compétentes à 4°C (env 100µL par falcon).
12. Transférer les cellules electrocompétentes dans un eppendorf préalablement mis à 4°C et mélanger 50ng de mini-targetting cassette.
13. Transformer par électroporation. Transférer les cellules dans 1 mL de LB puis incubé à 30°C pdt 1h.
14. Etaler les bactéries (1 µL 100µL et le reste) sur le milieu de sélection LB/Kanamycine de l'électroporation. En se mettant sur Kanamycine on sélectionne les clones qui auront recombiné. La sélection Kanamycine est apportée par la mini-targetting cassette.
15. Incuber à 30°C pour 16-24h.

Le lendemain :

Préculture de 6 clones dans 3mL de LB/Kan ON à 30°C.

Purification des minipreps

Cf protocole habituel

Vérification du Minitargeting Vector MTV

Définir une stratégie de digestion entre le GRP et le MTV.

Essayer d'avoir au max 3 bandes. Utiliser soit une simple ou une double digestion.

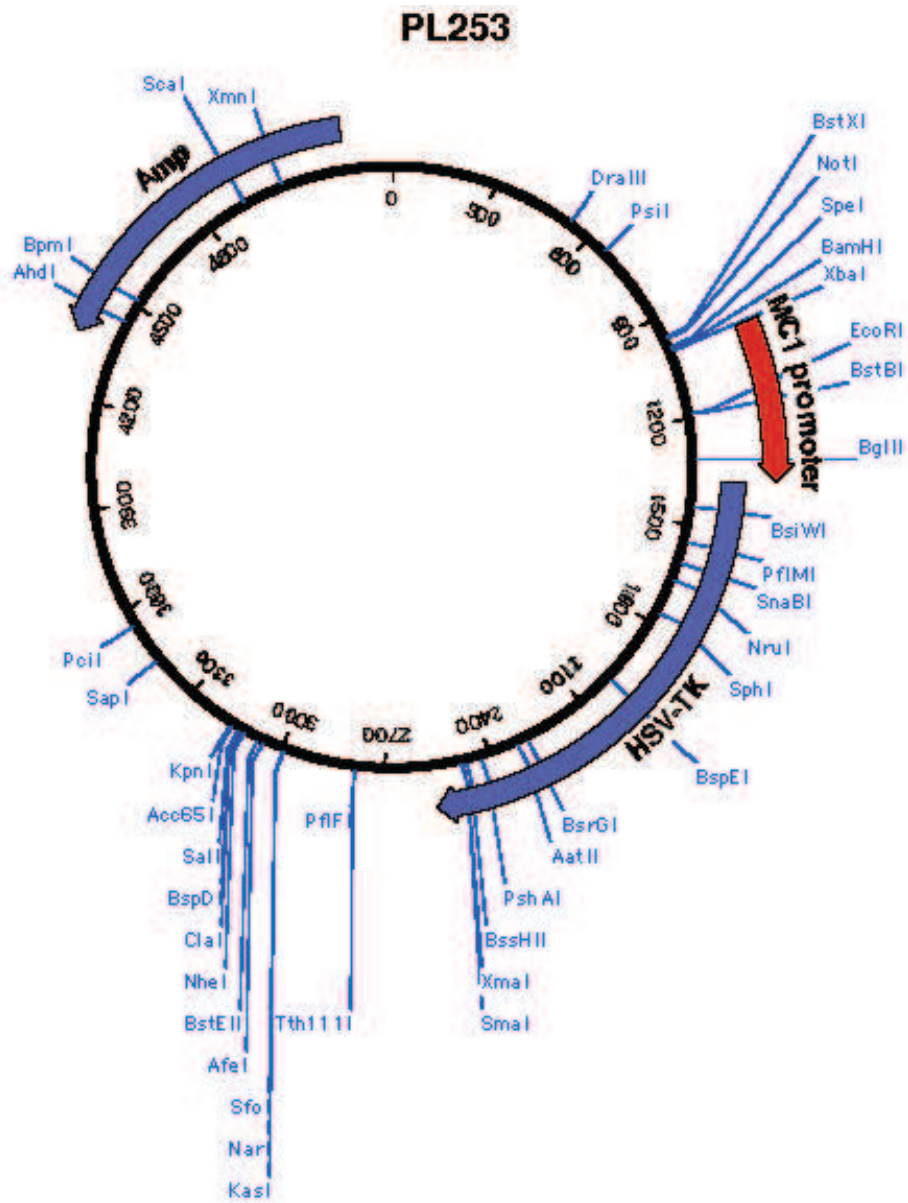
Idéalement on linéarise le GRP et on obtient 2 bandes pour le NTP.

Digérer 3µl d'ADN dans 20µl final.

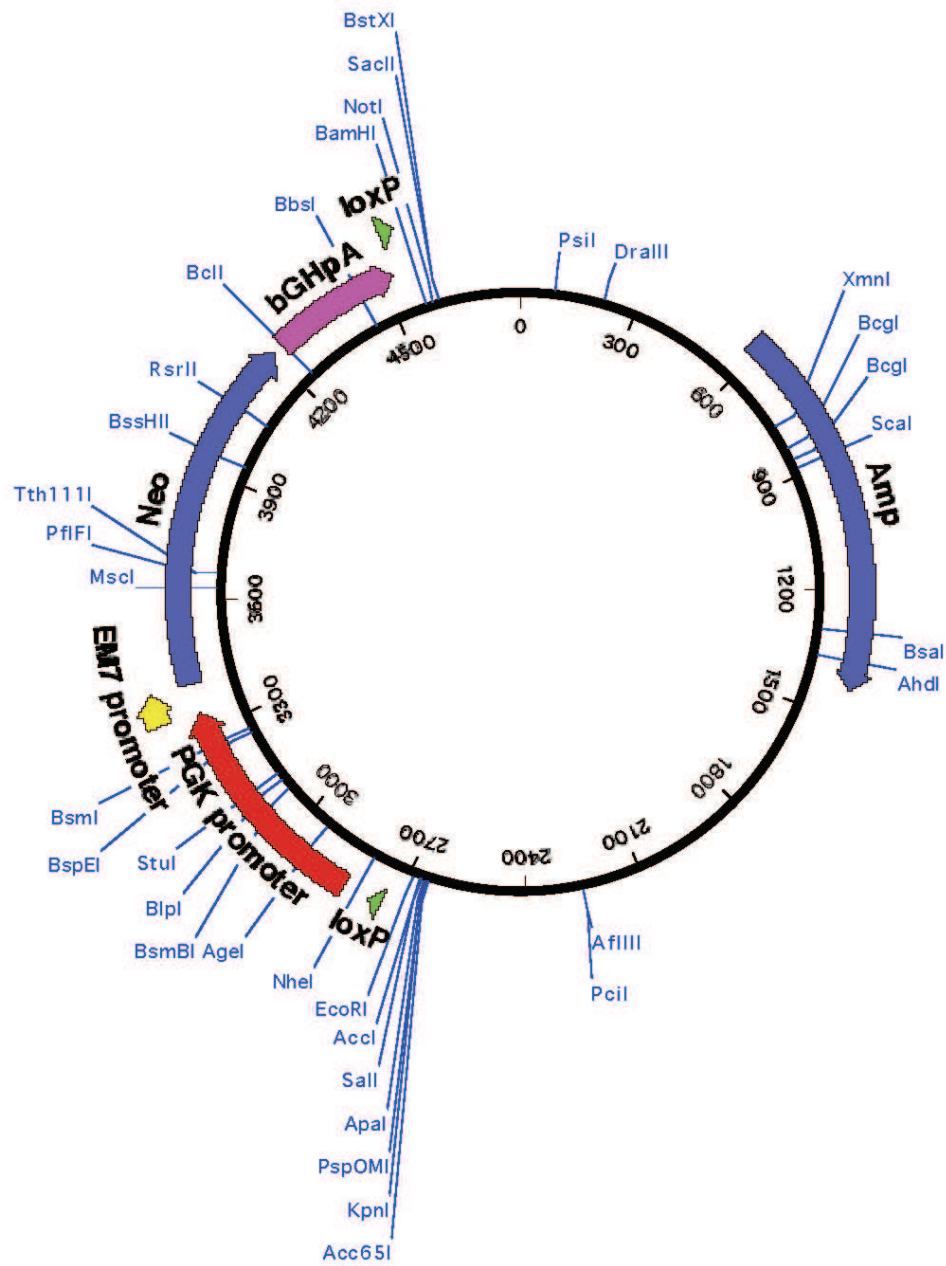
Mettre en collection le neo-targeted plasmid validé.







**PL452**



## Protocole de caryotypage

### *Jour 1 :*

1. Repartir  $2 \times 10^6$  cellules sur une boîte de 6cm gélatinée et laisser incuber ON dans l'étuve

### *Jour 2 :*

2. Incuber avec de la colcemide  $0,06\mu\text{g/ml}$  (stock à  $60\mu\text{g/ml} = 1000\text{X}$  dans le DMEN) durant 2 à 3 heures.
3. **Préparation des cellules :** Sortir :
  - Trypsine
  - D10
  - PBS 1X stérile
  - PBS 1X non stérile
  - KCl 0,56%
  - Solution de fixation (3 Volumes de Méthanol et 1 volume d'acide acétique glacial)
4. Laver la boîte de 6 délicatement au PBS
5. Ajouter 1ml de trypsine
6. Incuber 5min (min) à  $37^\circ\text{C}$
7. Ajouter 4ml de D10
8. Dissocier les cellules jusqu'à l'obtention d'une suspension de cellules bien individualisées
9. Centrifuger 1000rpm, 5min
10. Pendant ce temps, préparer la solution de fixation
11. Enlever le milieu et laisser 200 $\mu\text{l}$ , resuspendre très délicatement le culot à l'aide d'une p200
12. Ajouter 5ml de KCl 0,56% et laisser 6 min à température ambiante (ceci fait gonfler les cellules.
13. Centrifuger 1000rpm, 5min
14. Enlever le surnageant, laisser 200 $\mu\text{l}$ , resuspendre très délicatement le culot à l'aide d'une p200
15. Resuspendre dans 5ml de solution de fixation. Incuber 20min à température ambiante
16. Pendant ce temps, préparer les lames
17. Enlever le surnageant, laisser 200 $\mu\text{l}$ , resuspendre très délicatement le culot à l'aide d'une p200
18. Resuspendre dans 5ml de solution de fixation fraîche. Incuber 5min à température ambiante
19. Enlever le surnageant, laisser 50 $\mu\text{l}$
20. Ajouter 500 $\mu\text{l}$  et resuspendre très délicatement le culot à l'aide d'une p200
21. **Préparation des lames :** Laver les lames dans l'acétone
22. Puis laver dans l' $\text{H}_2\text{O}$

23. Laisser sécher sur la plaque chauffante ~ 45°C
24. Immerger les lames dans un mélange Ethanol/H<sub>2</sub>O (20/30)
25. Enlever les lames et enlever l'excès de liquide sans secouer.
26. **Montage des cellules et des lames :** Disposer les lames à 45° d'inclinaison et faire tomber 2-3 gouttes de la suspension cellulaire, se placer environ à 40-50cm au-dessus de la lame
27. Laisser sécher sur la plaque chauffante
28. Laver les lames dans le PBS 1X (non stérile) pour enlever les résidus acides.
29. Sur la paillasse propre disposer quelques dizaines de cm de parafilm.
30. Y déposer autant de lamelles que de lames
31. Préparer DAPI 1X (1ul dans 1ml d'H<sub>2</sub>O)
32. Déposer 3 à 4 gouttes de Dapi sur chaque lamelle avec une p200
33. Retourner les lames pour que la face avec les cellules soit en contact avec les gouttes de Dapi.
34. Laisser 2min à l'obscurité (couvrir avec de l'Aluminium)
35. Agiter vigoureusement les lames dans un b cher de PBS 1X. La lamelle se décolle.
36. Laisser sécher le moins de temps possible.
37. Sur la paillasse propre disposer quelques dizaines de cm de parafilm et y déposer autant de lamelles que de lames
38. Mettre 2 à 3 gouttes de liquide de montage vectashield 4°C sur chaque lamelle
39. Retourner les lames sur les lamelles, s'assurer de la bonne fixation
40. Laisser sécher 1 nuit à l'obscurité dans un tiroir

### *Jour 3 :*

41. Finir par fixer au vernis si possible aux 4 angles (important pour l'observation au microscope)
42. Compter une dizaine de figures mitotiques.

## Immunoprécipitation de la chromatine

### *La veille :*

Le matin, ensemercer  $8.10^6$  cellules par boîte de 10cm gélatinée. Max 4\*10 boîtes d'un coup. Normalement le lendemain on a environ  $1,5.10^7$  cellules par boîte.

### *Jour 1 :*

Changer le milieu 2h avant.

#### **Préparation de la chromatine**

1. Crosslinker par ajout de 0.4% formaldéhyde dans le milieu de culture {110  $\mu$ l formaldéhyde 36,5% dans 10 ml de milieu de culture}. Incuber 10 min à température ambiante.
2. Arrêter la réaction de crosslink par ajout de glycine à une concentration finale de 0.125 M. Incuber 10 min à température ambiante. (stock 1,25M, préparé le jour même. 1 ml par boîte)
3. Rincer deux fois les cellules avec du PBS froid {8 ml}.
4. Scraper les cellules avec 1 à 2 ml de PBS contenant des inhibiteurs de protéases (PMSF+Complete) et transférer dans un falcon50. Conserver à 4°C.
5. Centrifuger les cellules 5 min à 2500 rpm (4°C).
6. Jeter le surnageant et ajouter 1 ml par boîte de FA/SDS + PMSF + Complete. {aliquoter par 2 ml dans des tubes de 2ml}. Incuber dans la glace 15 min.
7. Purifier la chromatine par centrifugation pendant 20 min à 12000 rpm. Jeter le surnageant.
8. Homogénéiser la chromatine et incuber 1hr dans 1 ml FA/SDS + PMSF + Complete sur une roue à 4°C.
9. Centrifuger la chromatine 10 min à 12000 rpm (4°C).
10. Regrouper tous les tubes. Resuspendre la chromatine dans 1,6 ml de FA/SDS + PMSF + Complete par 5 boîtes dans des tubes de 2 ml. Il est possible de s'arrêter là en conservant la chromatine à -80°C.
11. Procéder à la sonication de la chromatine. Sur l'ancien sonicateur 6 cycles de 40sec ON/20sec OFF, pulse, 60%, puissance 4. Alternative (attention! Pas très reproductible!): sur le Diagenode Bioruptor 17 cycles de 20 sec ON / 40 sec OFF (attention: volume maximal 500 $\mu$ l dans tubes de 1,5ml) Refroidir le bac au préalable.
12. Centrifuger 10 min à 14000 rpm et récupérer le surnageant. Attention: le culot doit être petit et dense (parfois un peu brun). Si le culot est trop gros, la chromatine n'aura pas été suffisamment bien soniquée. Conserver un aliquot de 20  $\mu$ l (pour vérifier la sonication et estimer la concentration de la chromatine), et conserver le reste de la chromatine en aliquot de 500  $\mu$ l à -80°C.

### **Vérification de la sonication et de la concentration de la chromatine**

1. Réverser le crosslink de l'aliquote de 20 µl. Ajouter 480 µl d'eau, 20 µl de NaCl 5M et 15µg de Rnase A. Incuber 4 heures ou overnight à 65°C.
2. Ajouter 10 µl d'EDTA 0.5M pH 8.0, 20 µl de Tris 1M pH 8.0, 2 µl de proteinase K 20 mg/ml et incuber 1 heure à 50°C.
3. Extraction au phénol chloroforme (ou QIAGEN PCR Purification kit).
4. Précipiter l'ADN avec 1/10 vol NaAc 3M, 2-2.5 vol EtOH 100% et 10 µg Glycogène (1 µl). Incuber 30 min at -20°C puis, centrifuger 15 min à vitesse maximale. Laver le culot avec EtOH 70% et centrifuger 10 min à vitesse maximale. Sécher le culot et resuspendre dans 50 µl H<sub>2</sub>O.
5. Estimer la quantité et la pureté de l'AND au nanodrop.
6. Faire migrer sur gel d'agarose 1.5% un échantillon de 5 µl.

Si la sonication est incomplète, réaliser des cycles de sonication supplémentaires.

## *Jour 2 :*

### **Preparation des billes couplées aux anticorps**

1. 50µl de Dynabeads IgG mouse par IP
2. Laver 3X 500µl H<sub>2</sub>O
3. Laver 2X 500µl PBS/BSA 0,1%
4. Resuspendre les billes dans 100µl PBS/BSA 0,1%
5. Ajouter 3µg anticorps HA ou Flag et incuber 1hr à 1300 rpm à 30°C

### **Blocage des billes**

Bloquer les billes dans 1 ml de FA/SDS + 0.5mg/ml de BSA et 0.2mg/ml yeast tRNA. Incuber 1 heure à 21°C.

(50 µl BSA 10 mg/ml et 8.76 µl ssDNA 4.38 mg/ml dans 1 ml FA/SDS).

### **Immunoprecipitation**

1. Pour chaque IP, préparer un tube de 1,5 ml avec 80-250 µg de chromatine soniquée (décongelée par centrifugation à 12000 rpm à 4°C) et compléter à 500 µl avec FA/SDS.
2. Ajouter les billes et incuber 2 heures à 21°C.
3. Jeter le surnageant.
4. Laver les billes avec 1 ml des tampons suivants:  
Low Salt Buffer, 1 fois  
High Salt Buffer, 1 fois  
Low LiCl, 1 fois  
TE, 2 fois
5. Jeter le surnageant.
6. Eluer avec 250 µl d'Elution Buffer (préparé le jour même, incubé à 65°C). Incuber 15 min à 65°C.
7. Collecter l'éluat.
8. Répéter l'atape d'élution et combiner les deux éluats dans un meme tube.
- 9.

**Reversion du crosslink**

Ajouter 20 µl NaCl 5M et 15µg Rnase A aux 500 µl d'éluat et incuber overnight à 65°C.

Ne pas oublier de reverser le crosslink d'un échantillon "input" (20 µl input + 480 µl Elution Buffer + 20 µl NaCl 5M et 20µg Rnase A => ON à 65°C).

*Jour3 :*

**Digestion à la Proteinase K**

Ajouter 10 µl EDTA 0.5M pH8, 20 µl Tris 1M pH8 et 2 µl proteinase K aux 500 µl de chromatine immunoprécipité et à l'échantillon "Input", et incuber 1 heure à 50°C.

**Purification de l'ADN**

1. Extraction au phénol-chloroforme
2. Précipiter à l'EtOH avec 1 µl Glycogène et resuspendre dans 50 µl H<sub>2</sub>O.



## Solutions

### **FA/SDS**

10 mM EDTA pH 8.0  
50 mM Tris pH 7.5  
150mM Nacl  
1mM EDTA  
1% Triton X100  
0,1% Na-Deoxycholate  
0,1%SDS

### **Low Salt Buffer**

0.1 % SDS  
1 % Triton X-100  
2 mM EDTA pH 8  
20 mM Tris, pH 8  
150 mM NaCl

(50 ml)  
500 µl SDS 10 %  
500 µl Triton X-100  
400 µl EDTA 250mM, pH 8  
1 ml Tris 1 M, pH 8  
1.5 ml NaCl 5 M

### **High Salt Buffer**

0.1 % SDS  
1 % Triton X-100  
2 mM EDTA pH 8  
20 mM Tris, pH 8  
500 mM NaCl

(50 ml)  
500 µl SDS 10 %  
500 µl Triton X-100  
400 µl EDTA 250mM, pH 8  
1 ml Tris 1 M  
5 ml NaCl 5 M

### **Low LiCl Buffer**

0.25 M LiCl  
1 % NP-40/Igepal  
1 % deoxycholate  
10 mM Tris, pH 8  
1 mM EDTA, pH 8

(50 ml)  
12.5 ml LiCl 1 M  
0.5 ml NP-40/Igepal  
5 ml deoxycholate 10 %  
0.5 ml Tris 1M, pH 8  
200 µl EDTA 250 mM, pH 8

### **IP elution Buffer**

1 % SDS  
0.1 M NaHCO<sub>3</sub>

(50 ml)  
5 ml SDS 10 %  
5 ml NaHCO<sub>3</sub> 1 M

## Analyse des résultats de ChIP-on-chip (BioConductor)

NB: Tout le texte en bleu correspond à du code R.

La méthode de **Chromatin ImmunoPrecipitation on Chip** (abrégée ChIP-on-chip) permet l'étude des protéines interagissant avec la molécule d'ADN. Il s'agit d'une combinaison de la technique de Chromatin Immunoprécipitation avec la méthode des puces à ADN. Elle est en général utilisée pour repérer des sites de fixation des facteurs de transcription. Elle permet donc de localiser ces sites et d'étudier les séquences d'ADN correspondantes. Contrairement aux autres techniques classiques d'étude des interactions protéines – ADN, l'ADN utilisé ici est directement récupéré *in vivo*.

L'une des étapes critiques dans les expériences de ChIP/chip est l'analyse des résultats. Plusieurs méthodes existent pour la détermination des régions significativement enrichies. Nous utilisons le logiciel R avec essentiellement le package *limma* du projet BioConductor.

### A) Scan des lames / analyse de l'image (GenePix)

Pour chaque lame, GenePix produit un fichier tabulé de type *nomdufichier.grp* qui correspond à la traduction de l'intensité des pixels des spots en données numériques.

Structure d'un fichier GRP :

#### GPR Header

A sample GPR file header and a description of each entry are shown below:

Entry	Description
<b>ATF 1.0</b>	File type and version number.
<b>29 48</b>	Number of optional header records and number of data fields (columns).
<b>"Type=GenePix Results 3"</b>	Type of ATF file.
<b>"DateTime=2002/02/09 17:15:48"</b>	Date and time when the image was acquired.
<b>"Settings=C:\Genepix\Genepix.gps"</b>	The name of the settings file that was used for analysis.
<b>"GalFile=C:\Genepix\Demo.gal"</b>	The GenePix Array List file used to associate Names and IDs to each entry.

"PixelSize=10"	Resolution of each pixel in $\mu\text{m}$ .
"Wavelengths=635 532"	Installed laser excitation sources in nm.
"ImageFiles=C:\Genepix\demo.tif 0 C:\Genepix\Genepix.tif 1"	The name and path of the associated TIF file(s).
"NormalizationMethod=None"	The type of normalization method used, if applicable.
"NormalizationFactors=1 1"	The normalization factor applied to each channel.
"JpegImage=C:\Genepix\demo.jpg"	The name and path of the associated Jpeg image files.
"StdDev=Type 1"	The type of standard deviation calculation selected in the Options settings.
"RatioFormulation=W1/W2 (635/532)"	The ratio formulation of the ratio image, showing which image is numerator and which is denominator.
"Barcode=00331"	The barcode symbols read from the image.
"BackgroundSubtraction=LocalFeature"	The background subtraction method selected in the Options settings.
"ImageOrigin=0, 0"	The origin of the image relative to the scan area.
"JpegOrigin=390, 4320"	The origin of the Results JPEG image (the bounding box of the analysis Blocks) relative to the scan area origin.
"Creator=GenePix 4.1.1.4"	The version of the GenePix Pro software used to create the Results file.
"Scanner=GenePix 4000B [serial number]"	Type and serial number of scanner used to acquire the image.
"FocusPosition=0"	The focus position setting used to acquire the image, in microns.
"Temperature=19.6127"	The temperature of the scanner, in degrees C.
"LinesAveraged=1"	The line average setting used to acquire the image.

"Comment=hyb 2673"	User-entered file comment.
"PMTGain=500 600"	The PMT settings during acquisition.
"ScanPower=100 100"	The amount of laser transmission during acquisition.
"LaserPower=1 1"	The power of each laser, in volts.
"LaserOnTime=5 5"	The laser on-time for each laser, in minutes.
"Filters=<Empty> <Empty>"	Emission filters used during acquisition (GenePix 4100 and 4200 only.)
"ScanRegion=100,100,2000,2000"	The coordinate values of the scan region used during acquisition, in pixels.
"Supplier="	Header field supplied in GAL file.
<i>Data record column headings</i>	Column titles for each measurement (see below).
<i>Data Records</i>	Extracted data.

## GPR Data

The list below describes each column of data in the Results file.

Column Title	Description
<b>Block</b>	the block number of the feature.
<b>Column</b>	the column number of the feature.
<b>Row</b>	the row number of the feature.
<b>Name</b>	the name of the feature derived from the Array List (up to 40 characters long, contained in quotation marks).
<b>ID</b>	the unique identifier of the feature derived from the Array List (up to 40 characters long, contained in quotation marks).
<b>X</b>	the X-coordinate in $\mu\text{m}$ of the center of the feature-indicator associated with the feature, where (0,0) is the top left of the image.
<b>Y</b>	the Y-coordinate in $\mu\text{m}$ of the center of the feature-indicator associated with the feature, where (0,0) is the top left of the image.

<b>Dia.</b>	the diameter in $\mu\text{m}$ of the feature-indicator.
<b>F635 Median</b>	median feature pixel intensity at wavelength #1 (635 nm).
<b>F635 Mean</b>	mean feature pixel intensity at wavelength #1 (635 nm).
<b>F635 SD</b>	the standard deviation of the feature pixel intensity at wavelength #1 (635 nm).
<b>B635 Median</b>	the median feature background intensity at wavelength #1 (635 nm).
<b>B635 Mean</b>	the mean feature background intensity at wavelength #1 (635 nm).
<b>B635 SD</b>	the standard deviation of the feature background intensity at wavelength #1 (635 nm).
<b>% &gt; B635 + 1 SD</b>	the percentage of feature pixels with intensities more than one standard deviation above the background pixel intensity, at wavelength #1 (635 nm).
<b>% &gt; B635 + 2 SD</b>	the percentage of feature pixels with intensities more than two standard deviations above the background pixel intensity, at wavelength #1 (635 nm).
<b>F635 % Sat.</b>	the percentage of feature pixels at wavelength #1 that are saturated.
<b>F532 Median</b>	median feature pixel intensity at wavelength #2 (532 nm).
<b>F532 Mean</b>	mean feature pixel intensity at wavelength #2 (532 nm).
<b>F532 SD</b>	the standard deviation of the feature intensity at wavelength #2 (532 nm).
<b>B532 Median</b>	the median feature background intensity at wavelength #2 (532 nm).
<b>B532 Mean</b>	the mean feature background intensity at wavelength #2 (532 nm).
<b>B532 SD</b>	the standard deviation of the feature background intensity at wavelength #2 (532 nm).
<b>% &gt; B532 + 1 SD</b>	the percentage of feature pixels with intensities more than one standard deviation above the background pixel intensity, at wavelength #2 (532 nm).
<b>% &gt; B532 + 2 SD</b>	the percentage of feature pixels with intensities more than two standard deviations above the background pixel intensity, at wavelength #2 (532 nm).
<b>F532 % Sat.</b>	the percentage of feature pixels at wavelength #2 that are saturated.
<b>Ratio of Medians</b>	the ratio of the median intensities of each feature for each wavelength, with the median background subtracted.
<b>Ratio of Means</b>	the ratio of the arithmetic mean intensities of each feature for each wavelength, with the median background subtracted.

<b>Median of Ratios</b>	the median of pixel-by-pixel ratios of pixel intensities, with the median background subtracted.
<b>Mean of Ratios</b>	the geometric mean of the pixel-by-pixel ratios of pixel intensities, with the median background subtracted.
<b>Ratios SD</b>	the geometric standard deviation of the pixel intensity ratios.
<b>Rgn Ratio</b>	the regression ratio of every pixel in a 2-feature-diameter circle around the center of the feature.
<b>Rgn R<sup>2</sup></b>	the coefficient of determination for the current regression value.
<b>F Pixels</b>	the total number of feature pixels.
<b>B Pixels</b>	the total number of background pixels.
<b>Sum of Medians</b>	the sum of the median intensities for each wavelength, with the median background subtracted.
<b>Sum of Means</b>	the sum of the arithmetic mean intensities for each wavelength, with the median background subtracted.
<b>Log Ratio</b>	log (base 2) transform of the ratio of the medians.
<b>Flags</b>	the type of flag associated with a feature.
<b>Normalize</b>	the normalization status of the feature (included/not included).
<b>F1 Median - B1</b>	the median feature pixel intensity at wavelength #1 with the median background subtracted.
<b>F2 Median - B2</b>	the median feature pixel intensity at wavelength #2 with the median background subtracted.
<b>F1 Mean - B1</b>	the mean feature pixel intensity at wavelength #1 with the median background subtracted.
<b>F2 Mean - B2</b>	the mean feature pixel intensity at wavelength #2 with the median background subtracted.
<b>SNR 1</b>	the signal-to-noise ratio at wavelength #1, defined by (Mean Foreground 1- Mean Background 1) / (Standard deviation of Background 1)
<b>F1 Total Intensity</b>	the sum of feature pixel intensities at wavelength #1
<b>Index</b>	the number of the feature as it occurs on the array.
<b>"User Defined"</b>	user-defined feature data read from the GAL file (GenePix Pro 4.1).

## B) Préliminaires

Après avoir installé R, télécharger BioConductor et les packages utilisés (la plupart des packages couramment utilisés sont déjà inclus dans BioConductor). Pour plus d'infos voir le site de BioConductor (<http://www.bioconductor.org/>).

Créer un dossier regroupant :

- les différents fichiers .gpr que vous souhaitez analyser en même temps
- le fichier .gal correspondant aux puces que vous avez utilisé
- le fichier TargetName (cf. ci-dessous) qui résume les expériences que vous souhaitez analyser.

## C) Importation des données dans R

Ouvrir R en double-cliquant sur l'icone. La console R apparait. Chaque début de ligne est indiqué par un ">", qui signifie que R attend vos instructions.

Tout d'abord, il faut vous placer dans le répertoire contenant les données à analyser. Dans le menu déroulant "Fichier", cliquer sur "Changer le répertoire courant...". La fenêtre qui s'ouvre vous permet de naviguer dans l'arborescence de l'ordinateur.

Méthode alternative : `>setwd('adresse')`. exemple : `>setwd('D:/ChIP on chip')`

Attention ! Dans Windows le chemin d'accès utilise des "/" et non des "\".

`> getwd()` permet de visualiser le répertoire de travail en cours.

Ensuite, soit on peut taper soi-même le code pas à pas, soit on peut exécuter un script déjà écrit. Pour la deuxième solution : dans le menu déroulant "Fichier", cliquer sur "Sourcer du code R...". Choisir le fichier contenant le code (ici il s'agit du fichier analyse\_LIMMA.R). Le script se lance et vous n'avez plus qu'à lire les instructions qui apparaissent à l'écran.

1ère étape : ouvrir les "packages" que l'on va utiliser

```
> library("Biobase")
> library("limma")
> library("marray")
> library("convert")
```

Le fichier "TargetName":

La première étape dans la plupart des analyses consiste à lire un fichier qui décrit quel échantillon a été hybridé sur quel puce, et donne le nom du fichier correspondant. Le fichier target est un fichier tabulé, qui peut être créé dans excel (enregistrer au format .txt).

Exemple :

```
> targets <- readTargets("TargetName.txt")
```

```
> targets
```

SlideNumber	FileName	Cy3	Cy5
1	1	lame1.gpr	IN IP WT
2	2	lame2.gpr	IN IP WT
3	3	lame4.gpr	IN IP WT
4	4	lame5.gpr	IN IP Rpb2
5	5	lame6.gpr	IN IP Rpb2
6	6	lame7.gpr	IN IP Rpb2
7	7	lame8.gpr	IN IP Rpb2
8	8	lame9.gpr	IN IP Rpb9
9	9	lame10.gpr	IN IP Rpb9
10	10	lame11.gpr	IN IP Rpb9
11	11	lame12.gpr	IN IP Rpb9
12	12	lame13.gpr	IN IP Rpb4
13	13	lame14.gpr	IN IP Rpb4
14	14	lame15.gpr	IN IP Rpb4
15	15	lame16.gpr	IN IP Rpb4

Lecture des fichiers:

La lecture des fichiers se fait grâce à la fonction `read.maimages()` :

```
> f <- function(x) as.numeric(x$Flags>-99)
> RG <- read.maimages(targets$FileName, source="genepix", wt.fun=f)
```

Signification de cette ligne : l'objet `RG` va stocker le résultat de l'exécution de `read.maimages()`, qui prend pour arguments :

- le ou les fichiers à traiter, ici : la colonne `FileName` du fichier `targets`
- le logiciel à l'origine des fichiers, ici : `genepix`
- la prise en compte des spots flaggés, ici on attribue aux spots flaggés (qui ont une valeur de -100 après le flagging dans `GenePix`) le poids 0, les spots non flaggés auront un poids égal à 1.

Qu'est ce qu'une `RGList`?

A simple list-based class for storing red and green channel foreground and background intensities for a batch of spotted microarrays. `RGList` objects are normally created by [read.maimages](#).

## Slots/List Components

`RGList` objects can be created by `new("RGList", RG)` where `RG` is a list. Objects of this class contains no slots (other than `.Data`), but objects should contain the following list components:

R:	numeric matrix containing the red (cy5) foreground intensities. Rows correspond to spots and columns to arrays.
G:	numeric matrix containing the green (cy3) foreground intensities. Rows correspond to spots and columns to arrays.



Optional components include

Rb:	numeric matrix containing the red (cy5) background intensities
Gb:	numeric matrix containing the green (cy3) background intensities
weights:	numeric matrix of same dimension as R containing relative spot quality weights. Elements should be non-negative.
other:	list containing numeric matrices of other spot-specific information.
genes:	data.frame containing probe information. Should have one row for each spot. May have any number of columns.
targets:	data.frame containing information on the target RNA samples. Rows correspond to arrays. May have any number of columns. Usually includes columns <code>cy3</code> and <code>cy5</code> specifying which RNA was hybridized to each array.
printer:	list containing information on the process used to print the spots on the arrays. See <a href="#">PrintLayout</a> .

Définition du nombre de lignes de l'objet RG:

```
> n <- length(RG$targets[,1])
```

```
> RG$genes <- readGAL()
```

Read a GenePix Array List (GAL) file into a dataframe.

```
> names(RG$genes)
```

Donne le nom des colonnes de la catégorie genes de la RGList.

```
> table(RG$genes$ControlType)
```

Donne sous forme de tableau le contenu de la catégorie ControlType de la catégorie genes de la RGList.

```
> r <- RG$genes$Row
```

```
> c <- RG$genes$Col
```

```
> nr <- max(r)
```

```
> nc <- max(c)
```

```
> RG$printer <- list("ngrid.c"=1, "ngrid.r"=1, "nspot.r"= nr, "nspot.c"=nc)
```

## D) Contrôle qualité

### 1. imageplot

```
> index <- (r-1)*nc+c
```

```
> for (i in 1:n){
```

```
> y <- rep(NA, nr*nc)
```

```
> filename <- paste("imageplot",i,".jpg",sep="")
```

```
> jpeg(filename = filename, quality = 100)
```

```
> y[index] <- RG[["Rb"]][,i]
```

```
> imageplot(y, layout=RG$printer)
```

```
> dev.off()
```

> }

## 2. MA plot

L'application du « MA plot » aux puces à ADN a été proposée par l'équipe de Speed T (NAR, 2002)

Une constatation: Le ratio R/G est souvent sous la dépendance de l'intensité

Pour permettre de mettre en exergue cet artefact

Calculer M (« minus » = le ratio)

$M = \text{Log}_2(R/G)$

(Si les données sont en logarithme,  $M = R - G$ )

Calculer A (« add » = la moyenne des valeurs logarithmique de R et G)

$M = \frac{1}{2} \text{Log}_2(R * G)$

(Si les données sont en logarithme,  $A = 1/2 (R + G)$ )

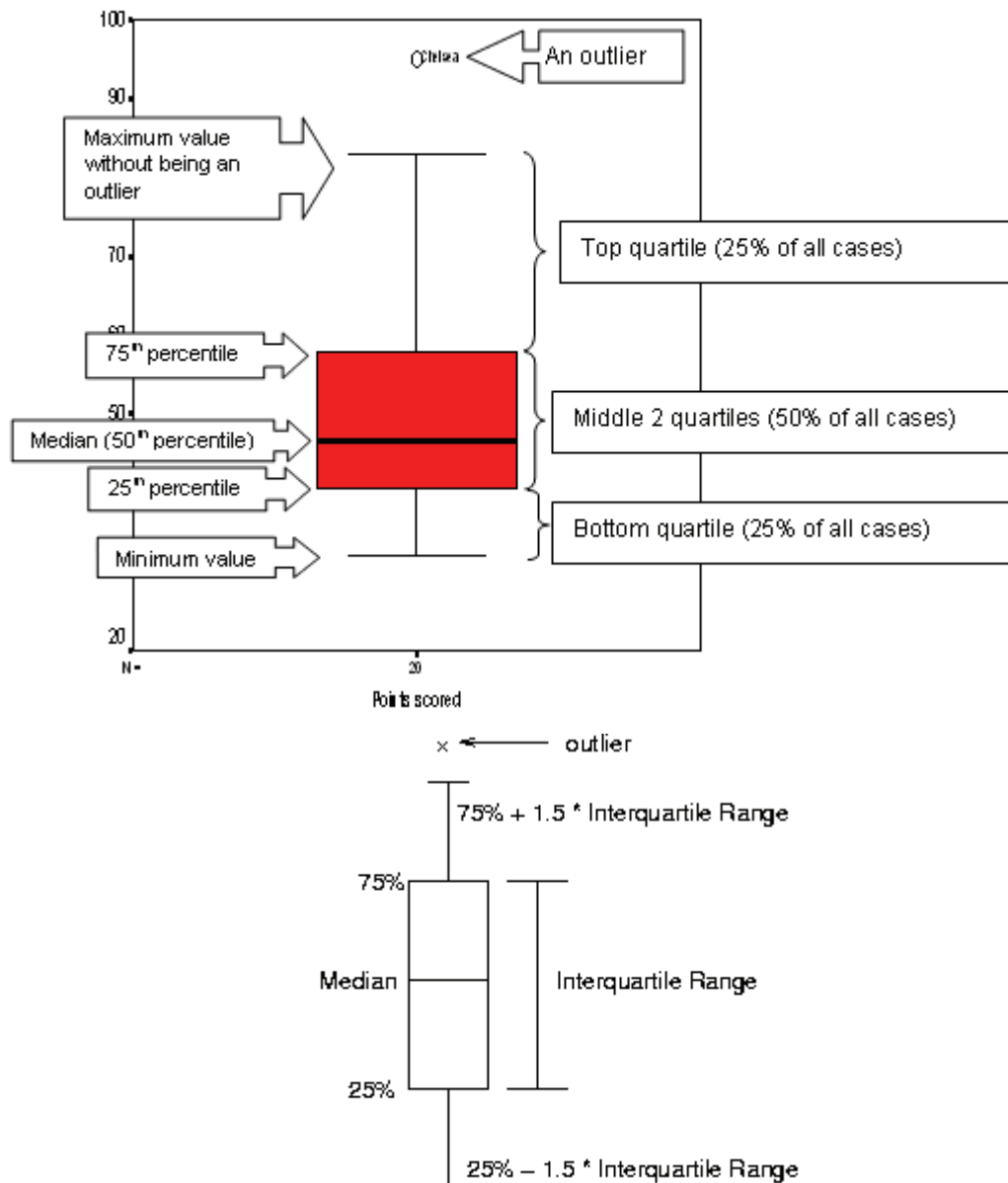
Représenter A en abscisse, M en ordonnée. M est souvent appelé log-ratio et A intensité moyenne.

Sous l'hypothèse que peu de fragments s'expriment différenciellement entre les deux conditions et que la quantité de fluorophore incorporée n'a pas d'influence sur le rapport, le nuage de points devrait se situer autour de l'axe des abscisses. Or nous observons une déformation qui est principalement due à la différence d'efficacité des fluorophores, et qui varie de plus d'une sonde à l'autre. Il existe différentes méthodes pour corriger cette déformation, la plus utilisée étant la méthode LOESS pour les données de transcriptome. (Attention! La normalisation Loess ne peut être appliquée pour les données de ChIP/chip, car d'une part le signal est uniquement positif (pas de up- ou down-régulation), et d'autre part la quantité de signal correspondant à une liaison peut être très importante).

```
> jpeg(filename = "MAplotbeforenormalization.jpg", quality = 100)
> par(mfrow=c(n,1), mar=rep(1,4))
> for(i in 1:n) {
> plotMA(RG[,i],xlab="A", ylab="M", status=RG$genes$ControlType, pch="*")
> }
> dev.off()
```

## 3. Boxplot (ou boîte à moustaches en français!)

Le boxplot constitue un résumé graphique de la distribution d'une variable quantitative, basé sur les quantiles. Il est formé d'une boîte limitée par le premier et le troisième quartile, divisé en 2 parties (inégaux pour une distribution asymétrique) par la médiane. Des tiges repèrent la position des quantiles les plus extrêmes. On représente également les valeurs (dites extrêmes) situées au-delà de ces limites (outlier). Le boxplot s'utilise en principe pour juger d'un coup d'œil de l'asymétrie d'une distribution et de l'existence de valeurs extrêmes.



```
> statsRb <- boxplot.stats(data.frame(RG$Rb))
> statsGb <- boxplot.stats(data.frame(RG$Gb))
> HingeBackground <- max(statsRb$stats[5], statsGb$stats[5]) * 1.05
> statsR <- boxplot.stats(data.frame(RG$R))
> statsG <- boxplot.stats(data.frame(RG$G))
> HingeSignal <- max(statsR$stats[5], statsG$stats[5]) * 1.05
> jpeg(filename = "boxplotbeforenormalization.jpg", quality = 100)
> par(mfrow = c(2, 2))
> boxplot(data.frame(RG$Rb), outline=FALSE, ylim=c(0, HingeBackground), main="Red
background", col="red")
> boxplot(data.frame(RG$Gb), outline=FALSE, ylim=c(0, HingeBackground), main="Green
Background", col="green")
> boxplot(data.frame(RG$R), outline=FALSE, ylim=c(0, HingeSignal), main="Red",
col="red")
> boxplot(data.frame(RG$G), outline=FALSE, ylim=c(0, HingeSignal), main="Green",
col="green")
> dev.off()
```

## E) Normalisation

- Pourquoi normaliser ?

Marquage par des fluorophores différents lors de réactions chimiques indépendantes

Les intensités de fluorescence sont mesurées par des lasers différents dans des longueurs d'onde différentes

### 1. Soustraction du bruit de fond

Il existe différentes manières de traiter le bruit de fond. Une approche simple consiste à soustraire à l'intensité mesurée au niveau des spots. Toutefois il est de plus en plus conseillé de ne pas soustraire le bruit de fond... (Tendance à accroître la variabilité du signal)

```
> RGsu <- backgroundCorrect(RG, method="subtract")

> jpeg(filename = "MAplotafterbgsubtraction.jpg", quality = 100)
> par(mfrow = c(2, 1))
> plotMA(RG, array=n, status=RG$genes$ControlType, pch="*", main="No background
correction")
> plotMA(RGsu, array=n, status=RG$genes$ControlType, pch="*", main="Background
subtraction")
> dev.off()
```

### 2. Normalize Within Arrays

Within array normalization adjusts array probe intensities so that values have consistent and meaningful interpretation across experiments. (For example, a log ratio  $\log_2$  (R/G) of 2 in a two-color experiment corresponds to a 4-fold enrichment measured in one channel (red) compared to another channel (green)).

Nous réalisons une normalisation par la médiane dite globale (global median location normalization). Cette méthode suppose que les intensités rouge et vert sont reliées par un facteur constant, et que le centre de la distribution est de zéro. Le log ratio de chaque spot est calculé, puis pour chaque puce, la médiane des log ratios est ensuite calculée. Cette valeur est alors soustraite au log ratio de chaque spot. La distribution de l'intensité des log ratios est ainsi "forcée" à avoir une médiane de 0 pour chaque puce.

```
> isGene=RGsu$genes$ControlType == "false"
> Mmedian <- normalizeWithinArrays(RGsu[isGene,], method="median",
controls=spots=NULL, bc.method="subtract", offset=0)

> jpeg(filename = "boxplotafternormalization.jpg", quality = 100)
> boxplot(data.frame(Mmedian$M), outline=FALSE, ylim=c(-6, 6), main="M median
normalized intensities", col="lightblue")
> dev.off()
> jpeg(filename = "densityplot.jpg", quality = 100)
> plotDensities(Mmedian)
> dev.off()
> jpeg(filename = "MAplotafternormalization.jpg", quality = 100)
```

```

> par(mfrow = c(3, 1))
> plotMA(RG, array=n, status=RG$genes$ControlType, pch="*", main="No background
correction")
> plotMA(RGsu, array=n, status=RG$genes$ControlType, pch="*", main="Background
substraction")
> plotMA(Mmedian, array=n, status=RG$genes$ControlType, pch="*", main="median
Normalisation")
> dev.off()

```

La normalisation génère une MAList

M: numeric matrix containing the M-values or log-2 expression ratios. Rows correspond to spots and columns to arrays.

A: numeric matrix containing the A-values or average log-2 expression values

### 3. Normalize Between Arrays (optionnel)

Between array normalization removes sources of variation of nonbiological origin among replicate arrays prior to combining them or comparing them for analysis.

La normalisation par la méthode des quantiles force les différents échantillons à avoir la même distribution des intensités (méthode proposée par Yang and Thorne (2003) pour les puces 2-couleurs).

```

> ANSWER <- readline("Do you want to perform a normalization between arrays? y/n ")
> if (ANSWER == "y") Mmedian2 <- normalizeBetweenArrays(Mmedian,
method="quantile")
> if (ANSWER == "n") Mmedian2 = Mmedian

> if (ANSWER == "y") {
> jpeg(filename = "densityplotafterban.jpg", quality = 100)
> plotDensities(Mmedian2)
> dev.off()
> jpeg(filename = "MAplotafterbanormalization.jpg", quality = 100)
> par(mfrow = c(3, 1))
> plotMA(RG, array=n, status=RG$genes$ControlType, pch="*", main="No background
correction")
> plotMA(RGsu, array=n, status=RG$genes$ControlType, pch="*", main="Background
substraction")
> plotMA(Mmedian, array=n, status=RG$genes$ControlType, pch="*", main="median
Normalisation")
dev.off()
> jpeg(filename = "boxplotafterbanormalization.jpg", quality = 100)
> boxplot(data.frame(Mmedian2$M), outline=FALSE, ylim=c(-6, 6), main="M median
normalized intensities", col="lightblue")
> dev.off()
> }

```

#### 4. Design matrix

```
> f=factor(targets$Cy5, levels=c(unique(targets$Cy5)))
> design=model.matrix(~0 + f)
```

#### 5. Modèle linéaire, test statistique, False discovery rate

This function fits multiple linear models. It accepts data from a experiment involving a series of microarrays with the same set of probes. A linear model is fitted to the expression data for each probe. The expression data should be log-ratios for two-color array platforms or log-expression values for one-channel platforms. The coefficients of the fitted models describe the differences between the RNA sources hybridized to the arrays.

`lmFit` has two main arguments, the expression data and the design matrix. The design matrix is essentially an indicator matrix which specifies which target RNA samples were applied to each channel on each array. There is considerable freedom in choosing the design matrix - there is always more than one choice which is correct provided it is interpreted correctly.

`lmFit` produces a fitted model object of class `MArrayLM` containing coefficients, standard errors and residual standard errors for each gene.

After fitting a linear model, the standard errors are moderated using a simple empirical Bayes model using `ebayes` or `eBayes`. A moderated t-statistic and a log-odds of differential expression is computed for each contrast for each gene. These functions is used to rank genes in order of evidence for differential expression. They use an empirical Bayes method to shrink the probe-wise sample variances towards a common value and to augmenting the degrees of freedom for the individual variances (Smyth, 2004). The empirical Bayes moderated t-statistics test each individual contrast equal to zero. For each probe (row), the moderated F-statistic tests whether all the contrasts are zero. The F-statistic is an overall test computed from the set of t-statistics for that probe. This is exactly analogous the relationship between t-tests and F-statistics in conventional anova, except that the residual mean squares and residual degrees of freedom have been moderated between probes.

`topTable` presents a list of the genes most likely to be differentially expressed for a given contrast. It summarizes the results of the linear model, performs hypothesis tests, and adjusts the p-values for multiple testing. The most popular form of adjustment is "fdr", which is Benjamini and Hochberg's method to control the false discovery rate. The meaning of "fdr" adjusted p-values is as follows. If all genes with p-value below a threshold, say 0.05, are selected as differentially expressed, then the expected proportion of false discoveries in the selected group is controlled to be less than the threshold value, in this case 5%. The result of `topTable` is a dataframe with a row for the `number` top genes and the following columns:

genelist	if genelist was included as input
logFC (ou f+nom de l'expérience)	estimate of the log2-fold-change corresponding to the effect or contrast
AveExpr	average log2-expression for the probe over all arrays and channels, same as <code>Amean</code> in the <code>MarrayLM</code> object
t	moderated t-statistic
P.Value	raw p-value
adj.P.Value	adjusted p-value or q-value
B	log odds that the gene is differentially expressed

Le False Discovery Rate (FDR) est l'espérance du taux de faux positifs. Contrôler le FDR au seuil  $\alpha$ , par exemple de 5%, permet d'affirmer qu'en moyenne, le taux de faux positifs est inférieur à 5%.

```
> fit <- lmFit(Mmedian2, design)
> fit <- eBayes(fit)
> l=length(Mmedian2$A[,1])
> FDR=topTable(fit,number=l)
> write.csv2(FDR,file="resultat.csv")
```

Les données sont fréquemment exprimées sous forme de ratio (R/G, C/V). La transformation logarithmique en base 2 permet de:

Passer d'une échelle non symétrique à une échelle symétrique (Facilite la manipulation des données)

D'obtenir une distribution proche d'une loi normale (critère pour de nombreux tests statistiques)

## En bref: les statistiques et l'analyse des puces à ADN

(Librement adapté de la thèse de doctorat de Nolween Le Meur  
[http://www.irisa.fr/symbiose/people/nlemeur/TheseDoctorat\\_NolwennLeMeur.pdf](http://www.irisa.fr/symbiose/people/nlemeur/TheseDoctorat_NolwennLeMeur.pdf))

**H<sub>0</sub>** (hypothèse nulle) : il n'existe pas de différence significative au risque  $\alpha$  de se tromper

**H<sub>1</sub>** (hypothèse alternative) : il existe une différence significative

**$\alpha$  Erreur de type I** ou risque de première espèce.

Probabilité d'obtenir des faux positifs, *i.e.* dire qu'un gène est différentiellement exprimé alors qu'il n'est pas différentiellement exprimé.

**$\beta$  Erreur de type II** ou risque de seconde espèce.

Probabilité d'obtenir des faux négatifs, *i.e.* dire qu'un gène n'est pas différentiellement exprimé alors qu'il l'est.

**Confiance (confidence)  $1 - \alpha$**

Probabilité de ne pas obtenir un résultat « faux positif », *i.e.* probabilité de conclure qu'un gène n'est pas différentiellement exprimé lorsqu'il ne l'est pas en réalité.

**Puissance (power)  $1 - \beta$**

Probabilité de ne pas obtenir un résultat « faux négatif », *i.e.* probabilité de conclure qu'un gène est différentiellement exprimé lorsqu'il l'est véritablement.

***p*-value**

Probabilité *p* d'obtenir une valeur de ratio plus surprenante que celle observée, si l'hypothèse nulle est vraie. Plus la *p*-value est petite, moins les variations observées sont dues au hasard.

***q*-value**

Probabilité *q* d'obtenir des faux positifs quand l'hypothèse nulle est vraie. Moins la *q* value est grande, moins le risque d'obtenir des faux positifs est élevé.

D'un point de vue statistique, pour mettre en évidence des gènes différentiellement exprimés, il faut réaliser un test d'hypothèses. Un test d'hypothèses est la vérification de la validité de l'hypothèse nulle sur les données observées grâce à un modèle probabiliste.

L'hypothèse nulle est qu'il n'existe pas de différence (pas d'effet biologique) significative au risque  $\alpha$  de se tromper. Dans ce cas, les variations observées sont liées à l'erreur de mesure. Soit la comparaison de 100 gènes dans 2 populations différentes et le seuil de signification  $\alpha=5\%$ , l'hypothèse H<sub>0</sub> est que l'expression de ces gènes ne soient pas significativement différents avec une confiance  $1 - \alpha = 95\%$ . La probabilité de dire que les



gènes sont différents (rejeter  $H_0$ ) alors qu'ils ne le sont pas est  $\alpha$ , *i.e.* sur les 100 gènes testés 5 peuvent être des faux positifs. La capacité de mettre en évidence un gène différentiellement exprimé est la puissance du test ou  $1-\beta$ .

Pour mettre en évidence des gènes différentiels, nous calculons, sous l'hypothèse nulle, la probabilité  $p$  d'obtenir une valeur de ratio plus « extrême » que celle observée. Si cette probabilité, encore appelée *p-value*, est très petite, l'événement est « surprenant » et  $H_0$  est rejetée. Par exemple, si la *p-value*=0,01 cela signifie que nous avons 1% de chance d'observer ce phénomène (un gène différentiel) au hasard. Le niveau d'expression d'un gène est déclaré significativement différentiel, lorsque la *p-value* est inférieure à un risque  $\alpha$  fixé. Le risque  $\alpha$  est également qualifié d'erreur de type I (ou risque de première espèce) et représente la probabilité de rejeter l'hypothèse nulle alors qu'elle est vraie. Ce type d'erreur correspond au fait de dire que des gènes sont différentiellement exprimés alors qu'ils ne le sont pas. Ce sont des faux positifs.

L'un des enjeux des tests statistiques est de minimiser le nombre de faux positifs, c'est-à-dire l'erreur de type I. Il existe deux grandes catégories de contrôle de l'erreur de type I : les approches FWER (*Family Wise Error Rate*) et FDR (*False Discovery Rate*). Le contrôle FWER mesure la probabilité  $p$  de faire une ou plusieurs erreurs de type I parmi l'ensemble des hypothèses testées ou *false positive rate*. L'approche FDR estime la proportion  $q$  des erreurs de type I parmi les hypothèses rejetées (*i.e.* les gènes estimés différentiellement exprimés). Par analogie avec la *p-value*, le résultat du contrôle FDR est parfois appelé *q-value*.

L'autre objectif des tests statistiques est de minimiser le nombre de faux négatifs, c'est-à-dire conserver une puissance statistique suffisante. La puissance d'une expérience dépend du nombre de replicats biologiques utilisés. Plus le nombre de replicats biologiques est important, plus la puissance de l'expérience augmente.

Afin de prendre en considération le risque d'augmenter le nombre de faux positifs par la multiplication des comparaisons, les résultats des tests statistiques doivent être corrigés. Le mode de contrôle de l'erreur de type I détermine le type de correction à appliquer. Les corrections de Bonferroni, Šidák, Holm ou Hochberg ajustent les *p-values* pour le contrôle FWER, tandis que les tests de Benjamini-Hochberg et Benjamini-Yekutieli ajustent les *p-values* pour le contrôle FDR.

Si dans certain cas le contrôle de FWER est utile, les corrections apportées entraînent généralement une perte de la puissance des tests statistiques. Par conséquent, il est sans doute

plus intéressant de contrôler le taux de faux positifs parmi les hypothèses rejetées (FDR) que de minimiser le taux de faux positifs parmi l'ensemble des tests effectués (FWER).

La mise en évidence des gènes "différentiels" nécessite l'utilisation de tests statistiques. Suivant les caractéristiques de données, plusieurs types de tests peuvent être utilisés. Nous utilisons généralement une inférence de type bayésienne. Le théorème de Bayes vise à calculer les probabilités *a posteriori* d'un événement en fonction des probabilités *a priori* de cet événement. *A priori* et *a posteriori* s'entendent par rapport à la connaissance d'une information. Plus précisément, le principe de base du théorème de Bayes est de calculer la probabilité *a posteriori* d'un événement A, sachant qu'un événement B s'est produit, en fonction de sa probabilité.

Dans le théorème de Bayes, l'évènement A, dont nous cherchons la probabilité *a posteriori*, peut être une hypothèse, un modèle ou un paramètre numérique telle la variance d'un échantillon. Ainsi, afin de palier aux limites des tests *t* classiques (normalité, nombre d'échantillons importants) dans la recherche de gènes différentiellement exprimés, certains auteurs proposent une estimation de la variance des échantillons par le théorème de Bayes.

Lönnstedt et Speed (2002) suggèrent un modèle mixte combinant une approche linéaire à une estimation de la variance par une méthode bayésienne paramétrique. Ils calculent ainsi la statistique B ou *log posterior odds ratios* qui correspond au ratio entre la probabilité qu'un gène donné soit différentiellement exprimé sur la probabilité que ce gène ne soit pas différentiellement exprimé. Une variante de cette solution, adaptée à la problématique des puces à ADN, a été développée par Smyth (2004) sous la forme d'une librairie R nommée *limma* (disponible sur le site du projet Bioconductor<sup>35</sup>).

## Résumé

Au cours de ma thèse, je me suis intéressée au facteur de transcription TFIIS, un facteur d'élongation de l'ARN polymérase II impliqué dans la stimulation de l'activité de clivage intrinsèque de cette enzyme. L'étude de la localisation globale du facteur TFIIS sur le génome de *Saccharomyces cerevisiae* par ChIP-on-chip a révélé que TFIIS, en plus d'être présent sur les gènes de classe II (ie. transcrits par l'ARN polymérase II), est également présent sur l'ensemble des transcrits de classe III (ie. transcrits par l'ARN polymérase III), suggérant ainsi un rôle jusqu'alors inconnu de cette protéine dans la transcription par l'ARN polymérase III. Des expériences de génomique, de génétique et de biochimie nous ont permis de montrer que TFIIS est un facteur de transcription de l'ARN polymérase III impliqué dans le choix du site d'initiation de la transcription. Par la suite, j'ai souhaité poursuivre cette étude chez *Mus musculus*, afin de déterminer, entre autres, si le rôle de TFIIS dans la transcription par l'ARN polymérase III est conservé chez les eucaryotes pluricellulaires. Des expériences préliminaires révèlent que Tcea1, l'une des isoformes de TFIIS chez *M. musculus*, serait présent sur quelques gènes de classe II et III. Lors de ma thèse, j'ai également pris part à deux autres projets en cours dans le laboratoire. L'un a porté sur le rôle du Médiateur, un complexe multiprotéique coactivateur de la transcription par l'ARN polymérase II, dans la mise en place du complexe de préinitiation via le recrutement du facteur TFIIF. Le second projet a permis de montrer que l'ARN polymérase II agit comme un senseur de la disponibilité en nucléosides triphosphates dans la cellule.

## Abstract

During my thesis I worked on the transcription factor TFIIS, which is responsible for the stimulation of the cleavage activity of arrested RNA polymerase II (Pol II). The genome-wide location analysis of TFIIS in *Saccharomyces cerevisiae* using ChIP-on-chip experiments revealed that TFIIS is located on many Pol II-transcribed genes, but also on all Pol III-transcribed genes, thus suggesting an unsuspected role of this protein in Pol III transcription. Further *in vivo* and *in vitro* experiments allowed us to propose that TFIIS is a Pol III transcription factor implicated in start site selection. I then decided to pursue this study in *Mus musculus*, in order to determine if the role of TFIIS in Pol III transcription is conserved in higher eukaryotes. Preliminary experiments revealed that Tcea1, one of the TFIIS isoforms in *M. musculus*, is located on some Pol II and Pol III transcribed genes. During my thesis, I have also been involved in two other projects. One of them focused on the role of the Mediator, a multiprotein coactivator complex of the RNA polymerase II, in the formation of the preinitiation complex by recruiting the TFIIF factor. The other project allowed us to demonstrate that the RNA polymerase II acts as a sensor of the nucleoside triphosphate availability in the cell.